US 20240150758A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2024/0150758 A1**

LUANGSAY et al. (43) **Pub. Date:** **May 9, 2024**

(57) **ABSTRACT**

The present invention relates to a FUBPI inhibitor for use in treatment of an HBV infection, in particular a chronic HBV infection. The invention in particular relates to the use of FUBPI inhibitors for destabilizing cccDNA, such as HBV cccDNA. The invention also relates to nucleic acid molecules, such as oligonucleotides including siRNA, shRNA and antisense oligonucleotides, which are complementary to FUBPI and capable of reducing a FUBPI mRNA. Also comprised in the present invention is a pharmaceutical composition and its use in the treatment and/or prevention of a HBV infection.

Specification includes a Sequence Listing.
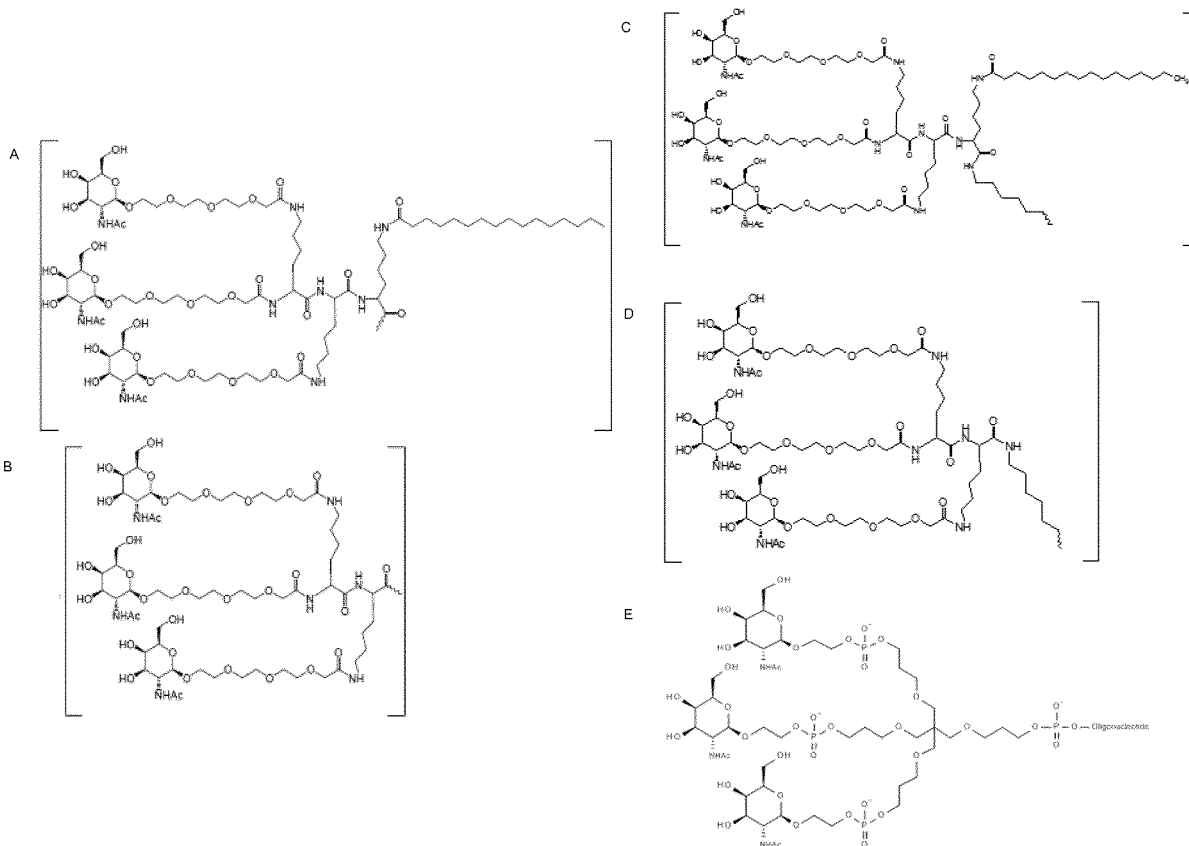
Figure 1

Figure 1 Cont'd

C

D

E

Figure 1 Cont'd

F

G

H
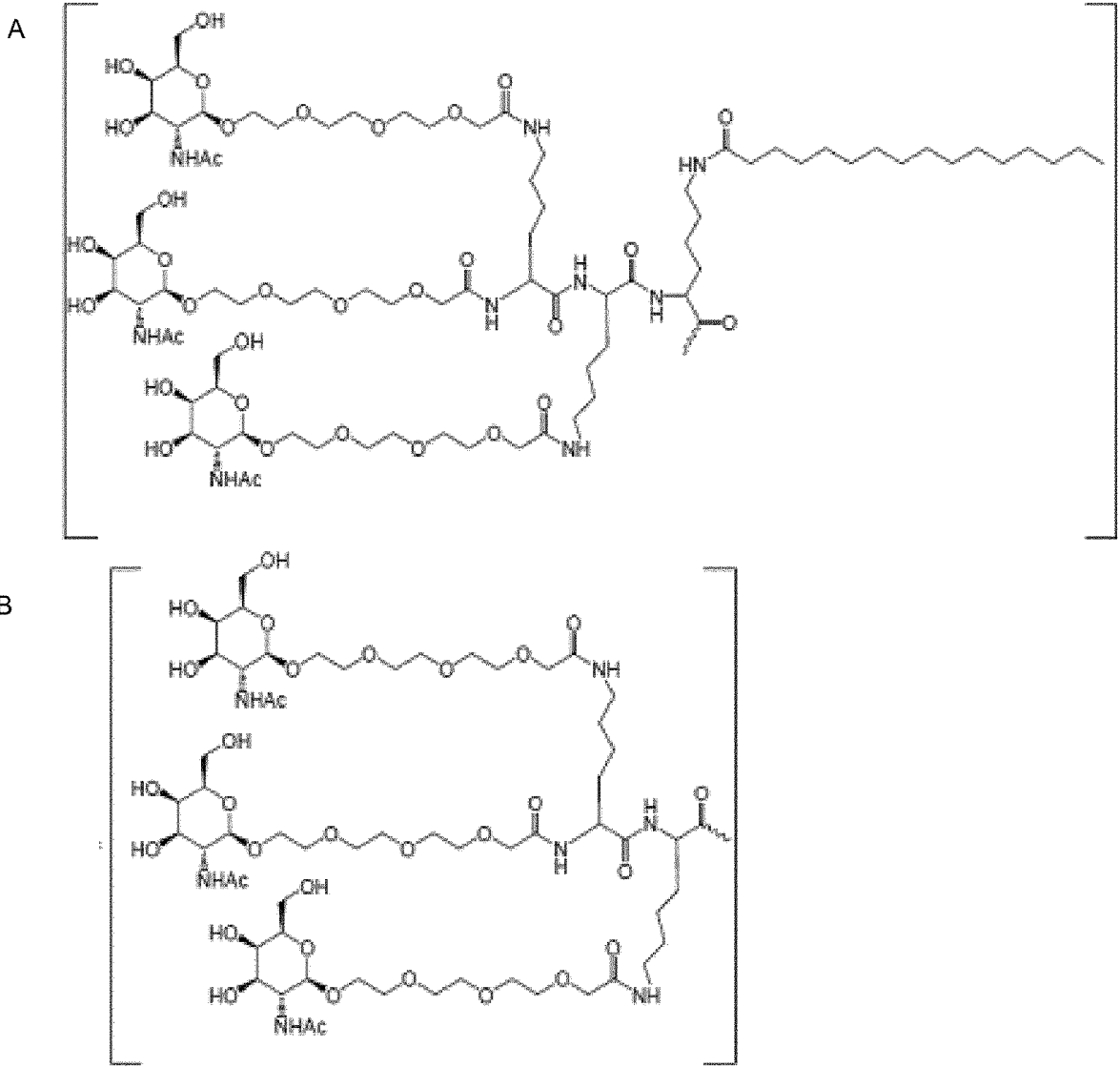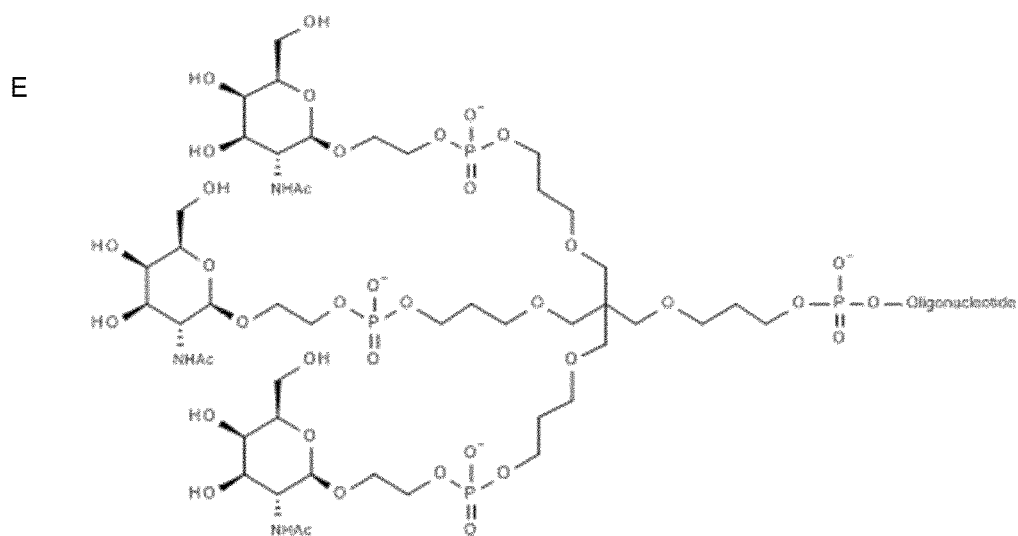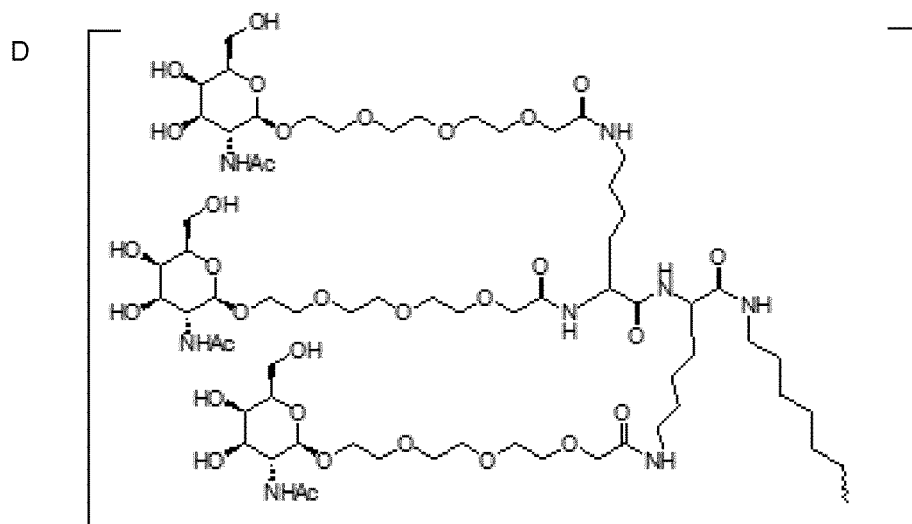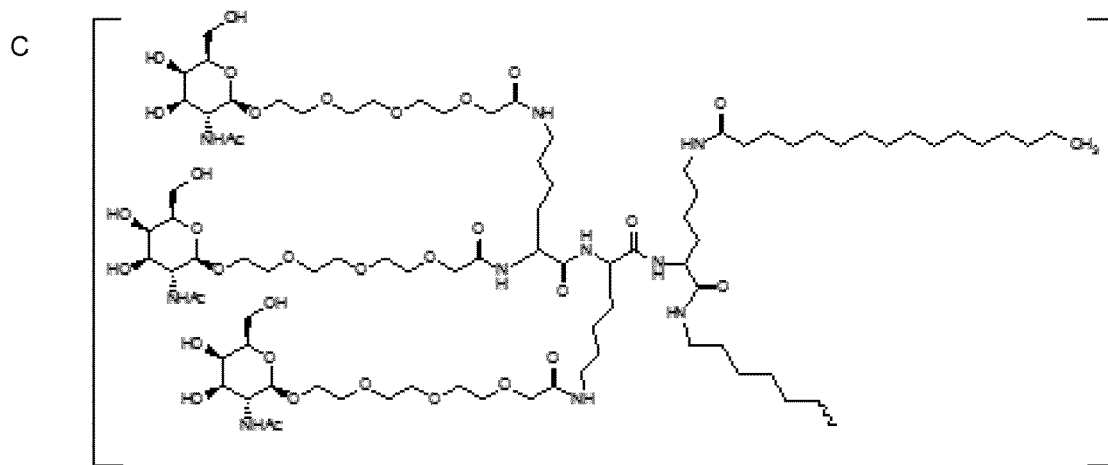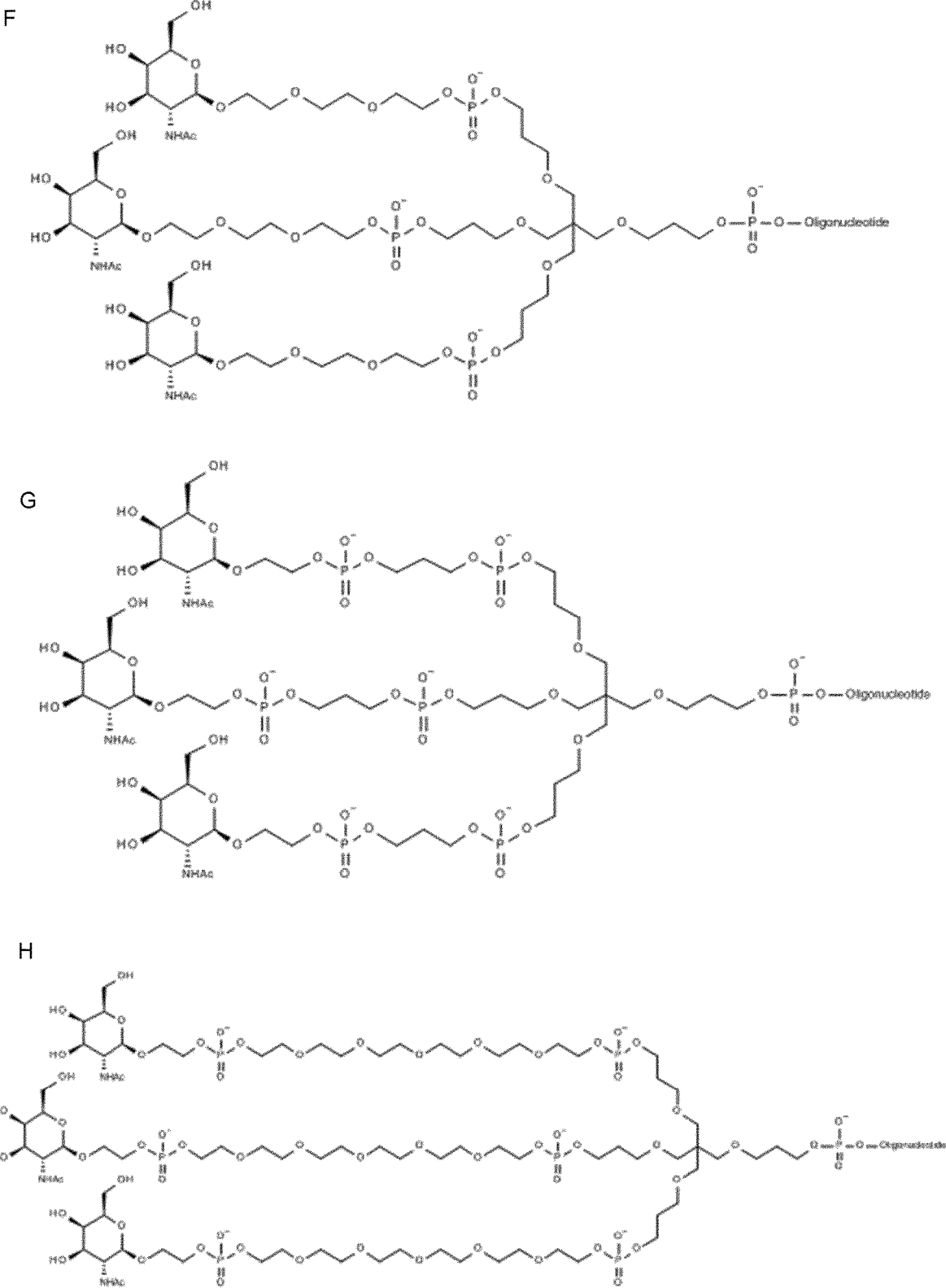
Figure 1 Cont'd
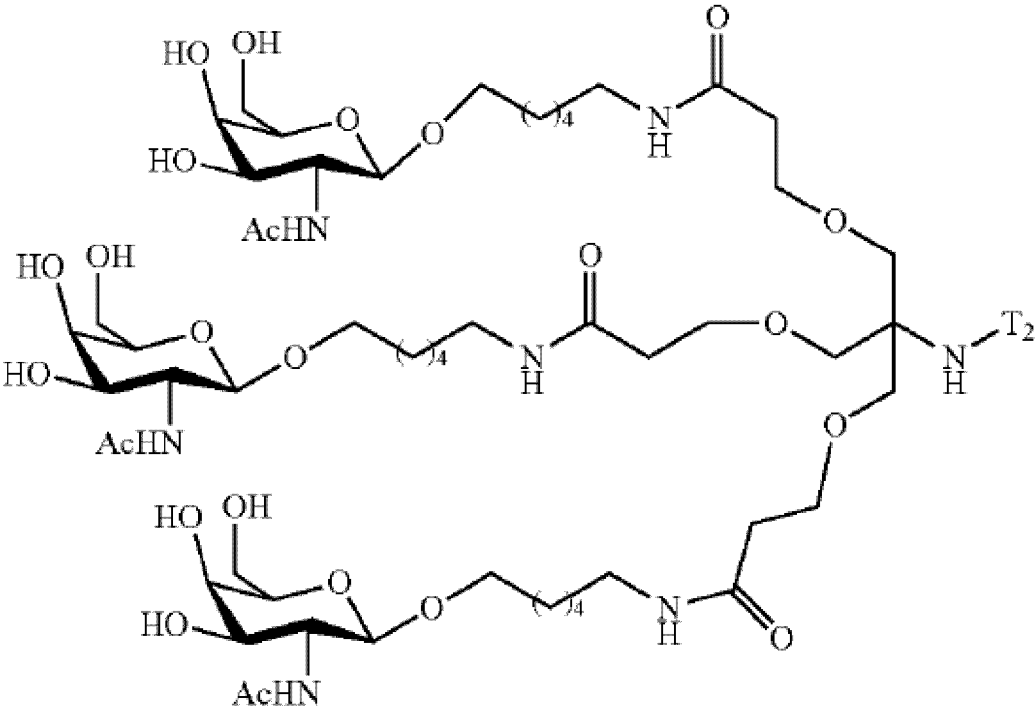
Figure 2
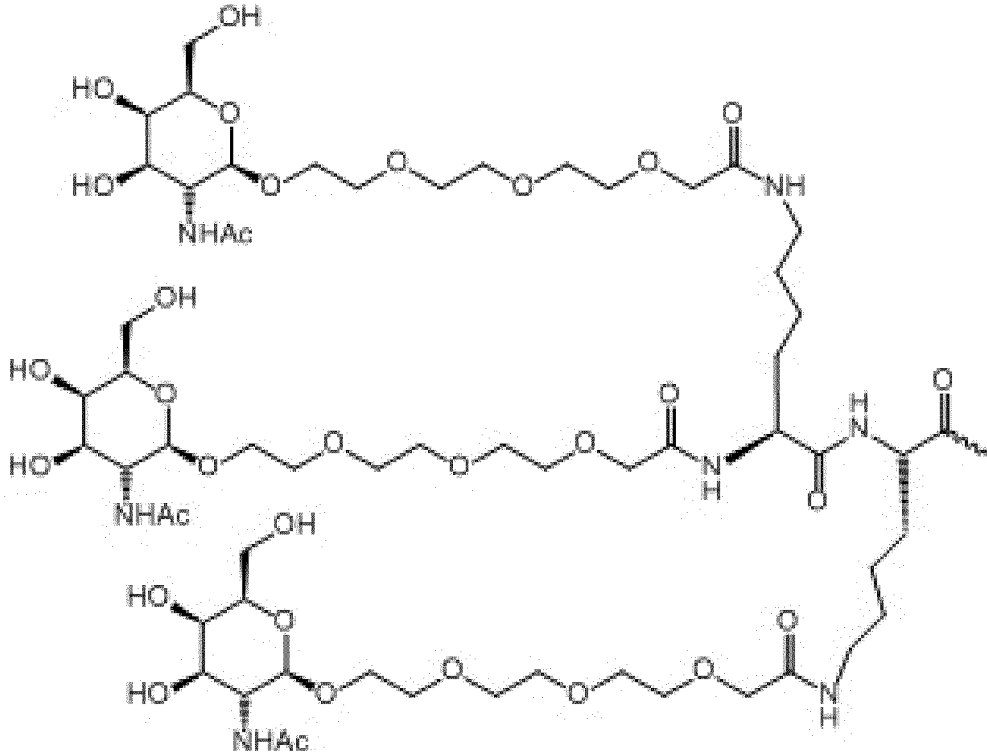
Figure 3
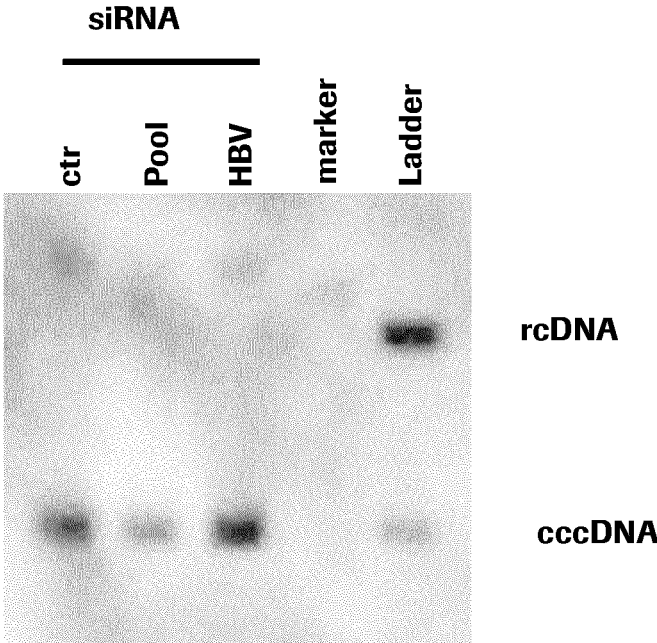
USE OF FUBP1 INHIBITORS FOR TREATING HEPATITIS B VIRUS INFECTION

FIELD OF THE INVENTION

[0001] The present invention relates to Far Upstream Element-Binding Protein 1 (FUBP1) inhibitors for use in treating and/or preventing a hepatitis B virus (HBV) infection, in particular a chronic HBV infection. The invention in particular relates to the use of FUBP1 inhibitors for destabilizing cccDNA, such as HBV cccDNA. The invention also relates to nucleic acid molecules, such as oligonucleotides including siRNA, shRNA and antisense oligonucleotides, which are complementary to FUBP1 and capable of reducing a FUBP1 target nucleic acid, such as mRNA. Also comprised in the present invention is a pharmaceutical composition and its use in the treatment and/or prevention of a HBV infection.

BACKGROUND

[0002] Far Upstream Element-Binding Protein 1 (FUBP1 or FBP1) is a single stranded DNA-binding protein that binds to multiple DNA elements. This protein is also thought to bind RNA and contains 3'-5' helicase activity with in vitro activity on both DNA-DNA and RNA-RNA duplexes. FUBP1 is known to activate the transcription of the proto-oncogene c-myc by binding to far upstream element (FUSE) located upstream of c-myc in undifferentiated cells. The protein is primarily present in the nucleus of the cell. Upregulation of FUBP1 has been observed in many types of cancers. Furthermore, FUBP1 can bind to and mediate replication of RNA from Hepatitis C virus and Enterovirus (Zhang and Chen 2013 Oncogene vol 32 p. 2907-2916).

[0003] FUBP1 has also been identified in Hepatocellular carcinoma (HCC) where it has been suggested to be involved in HCC tumorigenesis (Ramdzan et al 2008 Proteomics Vol 8 p. 5086-5096) and that FUBP1 is required for HCC tumour growth as illustrated using lentivirus expressed shRNA targeting FUBP1 (Rabenhorst et al 2009 Hepatology vol 50 p 1121-1129).

[0004] It has been demonstrated that knock down of FUBP1 with lentivirus expressed shRNA's enhances treatment response in ovarian cancer (Zhang et al 2017 Oncology Letters Vol 14 p. 5819-5824).

[0005] WO 2004/027061 disclose a screening method which involves the step of analyzing whether or not a test substance inhibits FBP and a medicinal composition for treating a proliferative disease which contains as the active ingredient(s) a substance inhibiting FBP.

[0006] Some small molecules inhibiting FUBP1 have been identified, all with the purpose of treating cancer (Huth et al 2004 J Med. Chen Vol 47 p. 4851-4857; Hauck et al 2016 Bioorganic & Medicinal Chemistry Vol 24 p. 5717-5729 Hosseini et al 2017 Biochemical Pharmacology Vol 146 p. 53-62 and Xiong et al 2016 Int J Onc vol 49 p 623). WO2004/017940 describes lipid based formulations of SN-38, it claims treatment of viral infection, in particular HIV, there is however no example supporting this.

[0007] Poly(U) Binding Splicing Factor 60 (PUF60) is a potentially regulator of both transcriptional and post-transcriptional steps of HBV pregenome expression. PUF60 is known to form a complex with FUBP1 in relation to c-myc repression. FUBP1 does, however, not participate in the PUF60 dependent regulation of HBV pregenome expression (Sun et al 2017 Scientific Reports 7:12874).

[0008] To our knowledge FUBP1 has never been shown to bind cccDNA or associated with cccDNA stability, nor has molecules inhibiting FUBP1 ever been suggested as cccDNA destabilizers or for the treatment of HBV infection.

[0009] HBV infection remains a major health problem worldwide which concerns an estimated 350 million chronic carriers. Approximately 25% of carriers die from chronic hepatitis, cirrhosis, or liver cancer. Hepatitis B virus is the second most significant carcinogen behind tobacco, causing from 60% to 80% of all primary liver cancer. HBV is 100 times more contagious than HIV.

[0010] The hepatitis B virus (HBV) is an enveloped, partially double-stranded DNA virus. The compact 3.2 kb HBV genome consists of four overlapping open reading frames (ORF), which encode for the core, polymerase (Pol), envelope and X-proteins. The Pol ORF is the longest and the envelope ORF is located within it, while the X and core ORFs overlap with the Pol ORF. The lifecycle of HBV has two main events: 1) generation of closed circular DNA (cccDNA) from relaxed circular (RC DNA), and 2) reverse transcription of pregenomic RNA (pgRNA) to produce RC DNA.

[0011] HBsAg quantification is a significant biomarker for prognosis and treatment response in chronic hepatitis B. However the achievement of HBsAg loss and seroconversion (functional cure) is rarely observed in chronically infected patients. Hepatitis B e-antigen (also called HBV envelope antigen or HBeAg) is a viral protein that is secreted by hepatitis B infected cells. HBeAg is associated with chronic hepatitis B infections and is used as a marker of active viral disease and a patient's degree of infectiousness.

[0012] Accordingly, reducing secretion of HBeAg in addition to secretion of HBsAg would lead to an improved inhibition of development of a chronic HBV infection as compared to the inhibition of secretion of HBsAg alone.

[0013] Current therapy such as Nucleos(t)ide analogues are molecules that inhibit HBV DNA synthesis but are not directed at reducing HBsAg level. Most therapies currently under development aim to reach a functional cure, with a durable HBsAg loss±anti-HBs seroconversion, undetectable serum DNA, and cccDNA in a transcriptionally inactive state, but do not address cccDNA persistence. These approaches do not therefore lead to a complete cure of HBV infection which is defined as durable HBV DNA and HBsAg loss with cccDNA elimination. The persistence of cccDNA in infected hepatocytes is the main barrier for eradicating the virus in CHB patients, and there is an urgent need to develop new therapies for the HBV complete cure that eliminates cccDNA.

OBJECTIVE OF THE INVENTION

[0014] The present invention shows that there is a correlation between the inhibition of FUBP1 and reduction of cccDNA in an HBV infected cell, which is relevant in the treatment of HBV infected individuals. An objective of the present invention is to identify FUBP1 inhibitors which reduce cccDNA in an HBV infected cell. Such FUBP1 inhibitors can be used in the treatment of HBV infection.

[0015] The present invention further identifies novel nucleic acid molecules, which are capable of inhibiting the expression of FUBP1 in vivo and in vitro.

BRIEF DESCRIPTION OF THE FIGURES

[0016] FIG. 1: Illustrates exemplary antisense oligonucleotide conjugates, where the oligonucleotide either is represented as a wavy line (A-D) or as "oligonucleotide" (E-H) or as T$_2$(I) and the asialoglycoprotein receptor targeting conjugate moieties are trivalent N-acetylgalactosamine moieties. Compounds A to D comprise a di-lysine brancher molecule, a PEG3 spacer and three terminal GalNAc carbohydrate moieties. In compound A and B the oligonucleotide is attached directly to the asialoglycoprotein receptor targeting conjugate moiety without a linker. In compound C and D the oligonucleotide is attached to the asialoglycoprotein receptor targeting conjugate moiety via a C6 linker. Compounds E-I comprise a commercially available trebler brancher molecule and spacers of varying length and structure and three terminal GalNAc carbohydrate moieties.

[0017] FIG. 2: Structural formula of the trivalent GalNAc cluster (GN2). GN2 is useful as conjugation moiety in the present invention. The wavy line illustrates the site of conjugation of the cluster to e.g. a C6 amino linker or directly to the oligonucleotide.

[0018] FIG. 3: Southern blot from HBV infected PPH cells treated either with control (ctr) or the pool of FUBP1 targeting siRNA's (Pool) or HBV siRNA (HBV).

SUMMARY OF THE INVENTION

[0019] One aspect of the present invention relates to a FUBP1 inhibitor for use in the treatment and/or prevention of Hepatitis B virus infection. In particular, a FUBP1 inhibitor capable of reducing cccDNA and/or pre-genomic RNA (pgRNA) is useful. In the event that the FUBP1 inhibitor is a small molecule it is advantageous if it interacts with the DNA binding domain of FUBP1 protein, and prevents or reduces FUBP1 binding to cccDNA. The FUBP1 inhibitor can also be a nucleic acid molecule of 12 to 60 nucleotides in length which is capable of reducing FUBP1 mRNA.

[0020] A further aspect of the present invention relates to nucleic acid molecules that inhibit expression and/or activity of FUBP1. In particular, a nucleic acid molecule of 12 to 60 nucleotides in length which comprises or consists of a contiguous nucleotide sequence of 12 to 30 nucleotides in length wherein the contiguous nucleotide sequence is at least 95% complementary to a mammalian FUBP1 target nucleic acid. Such nucleic acid molecules can be selected from single stranded antisense, siRNA and chemically produced shRNA molecules (not relying on cell based expression from plasmids or viruses).

[0021] A further aspect of the present invention relates to single stranded antisense oligonucleotides or siRNA that inhibit expression and/or activity of FUBP1. In particular, modified antisense oligonucleotides or modified siRNA comprising one or more 2' sugar modified nucleoside and one or more phosphorthioate linkage that reduce FUBP1 mRNA are of advantageous.

[0022] Further aspects of the invention are conjugates of nucleic acid molecules of the invention and pharmaceutical compositions comprising the molecules of the invention. In particular conjugates targeting the liver are of interest, such as GalNAc clusters.

Definitions

Nucleic Acid Molecule

[0023] The term "nucleic acid molecule" or "therapeutic nucleic acid molecule" as used herein is defined as it is generally understood by the skilled person, as a molecule comprising two or more covalently linked nucleosides (i.e. a nucleotide sequence). The nucleic acid molecule(s) referred to in the method of the invention are generally therapeutic oligonucleotides below 50 nucleotides in length. The nucleic acid molecules may be or comprise an antisense oligonucleotide, or may be another oligomeric nucleic acid molecule, such as a CRISPR RNA, a siRNA, shRNA, an aptamer, or a ribozyme. Therapeutic nucleic acid molecules are commonly made in the laboratory by solid-phase chemical synthesis followed by purification and isolation. shRNA's are however often delivered to cells using lentiviral vectors (see for example Soan and Yang 2010 N Am J Med Sci 2(12): 598) which are then transcribed to produce the single stranded RNA that will form a stem loop (hairpin) RNA structure that is capable of interacting with the RNA interference machinery (including the RNA-induced silencing complex (RISC)). When referring to a sequence of the nucleic acid molecule, reference is made to the sequence or order of nucleobase moieties, or modifications thereof, of the covalently linked nucleotides or nucleosides. The nucleic acid molecule of the invention is man-made, and is chemically synthesized, and is typically purified or isolated. In some embodiments the nucleic acid molecule of the invention is not a shRNA transcribed from a vector upon entry into the target cell. The nucleic acid molecule of the invention may comprise one or more modified nucleosides or nucleotides.

[0024] In some embodiments, the nucleic acid molecule of the invention comprises or consists of 12 to 60 nucleotides in length, such as from 13 to 50, such as from 14 to 40, such as from 15 to 30, such as from 16 to 22, such as from 16 to 18 or 15 to 17 contiguous nucleotides in length.

[0025] In some embodiments, the nucleic acid molecule or contiguous nucleotide sequence thereof comprises or consists of 24 or less nucleotides, such as 22, such as 20 or less nucleotides, such as 18 or less nucleotides, such as 14, 15, 16 or 17 nucleotides. It is to be understood that any range given herein includes the range endpoints. Accordingly, if a nucleic acid molecule is said to include from 12 to 30 nucleotides, both 12 and 30 nucleotides are included.

[0026] In some embodiments, the contiguous nucleotide sequence comprises or consists of 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 contiguous nucleotides in length

[0027] The nucleic acid molecule(s) are for modulating the expression of a target nucleic acid in a mammal. In some embodiments the nucleic acid molecules, such as for siRNAs, shRNAs and antisense oligonucleotides, are typically for inhibiting the expression of a target nucleic acid(s).

[0028] In one embodiment of the invention the nucleic acid molecule is selected from a RNAi agent, such as a siRNA or shRNA. In another embodiment the nucleic acid molecule is a single stranded antisense oligonucleotide, such as a high affinity modified antisense oligonucleotide interacting with RNaseH.

[0029] In some embodiments the nucleic acid molecule comprises phosphorothioate internucleoside linkages.

[0030] In some embodiments the nucleic acid molecule may be conjugated to non-nucleosidic moieties (conjugate moieties).

[0031] A library of nucleic acid molecules is to be understood as a collection of variant nucleic acid molecules. The purpose of the library of nucleic acid molecules can vary. In some embodiments, the library of nucleic acid molecules is composed of oligonucleotides with overlapping nucleobase sequence targeting one or more mammalian FUBP1 target nucleic acids with the purpose of identifying the most potent sequence within the library of nucleic acid molecules. In some embodiments, the library of nucleic acid molecules is a library of nucleic acid molecule design variants (child nucleic acid molecules) of a parent or ancestral nucleic acid molecule, wherein the nucleic acid molecule design variants retaining the core nucleobase sequence of the parent nucleic acid molecule.

Oligonucleotide

[0032] The term "oligonucleotide" as used herein is defined as it is generally understood by the skilled person as a molecule comprising two or more covalently linked nucleosides. Such covalently bound nucleosides may also be referred to as nucleic acid molecules or oligomers. Oligonucleotides are commonly made in the laboratory by solid-phase chemical synthesis followed by purification and isolation. When referring to a sequence of the oligonucleotide, reference is made to the sequence or order of nucleobase moieties, or modifications thereof, of the covalently linked nucleotides or nucleosides. The oligonucleotide of the invention is man-made, and is chemically synthesized, and is typically purified or isolated. The oligonucleotide of the invention may comprise one or more modified nucleosides or nucleotides, such as 2' sugar modified nucleosides.

Antisense Oligonucleotides

[0033] The term "Antisense oligonucleotide" as used herein is defined as oligonucleotides capable of modulating expression of a target gene by hybridizing to a target nucleic acid, in particular to a contiguous sequence on a target nucleic acid. The antisense oligonucleotides are not essentially double stranded and are therefore not siRNAs or shRNAs. Preferably, the antisense oligonucleotides of the present invention are single stranded. It is understood that single stranded oligonucleotides of the present invention can form hairpins or intermolecular duplex structures (duplex between two molecules of the same oligonucleotide), as long as the degree of intra or inter self complementarity is less than 50% across the full length of the oligonucleotide

[0034] Advantageously, the single stranded antisense oligonucleotide of the invention does not contain RNA nucleosides, since this will decrease nuclease resistance.

[0035] Advantageously, the antisense oligonucleotide of the invention comprises one or more modified nucleosides or nucleotides, such as 2' sugar modified nucleosides. Furthermore, it is advantageous that the nucleosides which are not modified are DNA nucleosides.

RNAi

[0036] Herein, the term "RNA interference (RNAi) molecule" refers to short double-stranded RNA molecule capable of inducing RNA-dependent gene silencing via the RNA-induced silencing complex (RISC) in a cell's cyto-plasm, where they interact with the catalytic RISC component argonaute. One type of RNAi molecule is a small interfering RNA (siRNA), which is a double-stranded RNA molecule composed of two complementary oligonucleotides, where the binding of one strand to complementary mRNA after transcription, leads to its degradation and loss of translation. A small hairpin RNA (shRNA) is a single stranded RNA molecule that forms a stem loop (hairpin) structure which upon expression is able to reduce mRNA via the DICER and RNA reducing silencing complex (RISC). RNAi molecules can be designed based on the sequence of the gene of interest (target nucleic acid). Corresponding RNAi can then be synthesized chemically or by in vitro transcription, or expressed from a vector or PCR product.

[0037] shRNA molecules are generally between 40 and 70 nucleotides in length, such as between 45 and 65 nucleotides in length, such as 50 and 60 nucleotides in length, and interacts with the endonuclease known as Dicer which is believed to processes dsRNA into 19-23 base pair short interfering RNAs with characteristic two base 3' overhangs which are then incorporated into an RNA-induced silencing complex (RISC). siRNA molecules are double stranded, with each strand being between 18 and 35 nucleotides in length, such as 20 to 30 nucleotides in length, such as 22 to 27 nucleotides in length. siRNA's are often designed with a two base 3' overhang to resemble the product produced by Dicer, which forms the RISC substrate. Effective extended forms of Dicer substrates have been described in U.S. Pat. No. 8,349,809 and U.S. Pat. No. 8,513,207, hereby incorporated by reference. Upon binding to the appropriate target mRNA, one or more endonucleases within the RISC cleave the target to induce silencing. RNAi olignucleotides may be chemically modified using modified internucleotide linkages and 2' sugar modified nucleosides, such as 2'-4' bicyclic ribose modified nucleosides, including LNA and cET or 2' substituted modifications like of 2'-O-alkyl-RNA, 2'-O-methyl-RNA, 2'-alkoxy-RNA, 2'-O-methoxyethyl-RNA (MOE), 2'-amino-DNA, 2'-fluoro-DNA, arabino nucleic acid (ANA), 2'-fluoro-ANA.

[0038] In some embodiments RNAi nucleic acid molecules comprise one or more phosphorothioate internucleoside linkages. In RNAi molecules phosphorothioate internucleoside linkages may reduce or the nuclease cleavage in RICS it is therefore advantageous that not al internucleoside linkages are modified. Phosphorothioate internucleoside linkages can advantageously be place in the 3' and/or 5' end of the RNAi nucleic acid molecule, in particular in the of the part of the molecule that is not complementary to the target nucleic acid (e.g. the sense stand or passenger strand in an siRNA molecule). The region of the RNAi molecule that is complementary to the target nucleic acid (e.g. the antisense or guide strand in an siRNA molecule) may however also be modified in the first 2 to 3 internucleoside linkages in the 3' and/or 5' terminal.

Contiguous Nucleotide Sequence

[0039] The term "contiguous nucleotide sequence" refers to the region of the oligonucleotide which is complementary to the target nucleic acid. The term is used interchangeably herein with the term "contiguous nucleobase sequence" and the term "oligonucleotide motif sequence". In some embodiments all the nucleotides of the oligonucleotide constitute the contiguous nucleotide sequence. In some embodiments the oligonucleotide comprises the contiguous nucleotide

sequence and may optionally comprise further nucleotide(s), for example a nucleotide linker region which may be used to attach a functional group to the contiguous nucleotide sequence. The nucleotide linker region may or may not be complementary to the target nucleic acid.

Nucleotides

[0040] Nucleotides are the building blocks of oligonucleotides and polynucleotides, and for the purposes of the present invention include both naturally occurring and non-naturally occurring nucleotides. In nature, nucleotides, such as DNA and RNA nucleotides comprise a ribose sugar moiety, a nucleobase moiety and one or more phosphate groups (which is absent in nucleosides). Nucleosides and nucleotides may also interchangeably be referred to as "units" or "monomers".

Modified Nucleoside

[0041] The term "modified nucleoside" or "nucleoside modification" as used herein refers to nucleosides modified as compared to the equivalent DNA or RNA nucleoside by the introduction of one or more modifications of the sugar moiety or the (nucleo)base moiety. In a preferred embodiment the modified nucleoside comprises a modified sugar moiety. The term modified nucleoside may also be used herein interchangeably with the term "nucleoside analogue" or modified "units" or modified "monomers". Nucleosides with an unmodified DNA or RNA sugar moiety are termed DNA or RNA nucleosides herein. Nucleosides with modifications in the base region of the DNA or RNA nucleoside are still generally termed DNA or RNA if they allow Watson Crick base pairing.

Modified Internucleoside Linkage

[0042] The term "modified internucleoside linkage" is defined as generally understood by the skilled person as linkages other than phosphodiester (PO) linkages, that covalently couples two nucleosides together. The nucleic acid molecules of the invention may therefore comprise modified internucleoside linkages. In some embodiments, the modified internucleoside linkage increases the nuclease resistance of the nucleic acid molecules of the invention compared to a phosphodiester linkage. For naturally occurring oligonucleotides, the internucleoside linkage includes phosphate groups creating a phosphodiester bond between adjacent nucleosides. Modified internucleoside linkages are particularly useful in stabilizing nucleic acid molecules, single stranded antisense as well as double stranded siRNA and shRNA molecules for in vivo use, and may serve to protect against nuclease cleavage at regions of DNA or RNA nucleosides in nucleic acid molecules of the invention, for example within the gap region of a gapmer oligonucleotide, as well as in regions of modified nucleosides.

[0043] In an embodiment, the nucleic acid molecule, e.g. antisense oligonucleotide, shRNA or siRNA, comprises one or more internucleoside linkages modified from the natural phosphodiester, such as one or more modified internucleoside linkages, that is for example more resistant to nuclease attack. Nuclease resistance may be determined by incubating the oligonucleotide in blood serum or by using a nuclease resistance assay (e.g. snake venom phosphodiesterase (SVPD), both are well known in the art. Internucleoside linkages which are capable of enhancing the nuclease

resistance of an oligonucleotide are referred to as nuclease resistant internucleoside linkages. In some embodiments at least 50% of the internucleoside linkages in the antisense oligonucleotide, or contiguous nucleotide sequence thereof, are modified, such as at least 60%, such as at least 70%, such as at least 75%, such as at least 80% or such as at least 90% of the internucleoside linkages in the oligonucleotide, or contiguous nucleotide sequence thereof, are modified. In some embodiments all of the internucleoside linkages of the oligonucleotide, or contiguous nucleotide sequence thereof, are modified. It will be recognized that, in some embodiments the nucleosides which link the oligonucleotide of the invention to a non-nucleotide functional group, such as a conjugate, may be phosphodiester. In some embodiments all of the internucleoside linkages of the oligonucleotide, or contiguous nucleotide sequence thereof, are nuclease resistant internucleoside linkages.

[0044] In some embodiments the internucleoside linkage comprises sulphur (S), such as a phosphorothioate internucleoside linkage. With the nucleic acid molecules of the invention it is advantageous to use phosphorothioate internucleoside linkages, in particular with single stranded antisense oligonucleotides.

[0045] Phosphorothioate internucleoside linkages are particularly useful due to nuclease resistance, beneficial pharmacokinetics and ease of manufacture. In some embodiments at least 20% of the internucleoside linkages in the nucleic acid molecule, or contiguous nucleotide sequence thereof, are phosphorothioate, such as at least 30%. For single stranded antisense oligonucleotides it is advantageous id at least 60% of the internucleoside linkages, such as at least 70%, such as at least 75%, such as at least 80% or such as at least 90% of the internucleoside linkages in the oligonucleotide, or contiguous nucleotide sequence thereof, are phosphorothioate. In some embodiments all of the internucleoside linkages of the antisense oligonucleotide, or contiguous nucleotide sequence thereof, are phosphorothioate. Nuclease resistant linkages, such as phosphorthioate linkages, are particularly useful in antisense oligonucleotide regions capable of recruiting nuclease when forming a duplex with the target nucleic acid, such as region G for gapmers. In an antisense gapmer oligonucleotide it is advantageous if the phosphodiester linkages, if present, are not located between contiguous DNA nucleosides in the gap region G. Advantageously, all the internucleoside linkages of the contiguous nucleotide sequence of the antisense oligonucleotide are phosphorothioate.

[0046] It is recognized that, as disclosed in EP 2 742 135, antisense oligonucleotides may comprise other internucleoside linkages (other than phosphodiester and phosphorothioate), for example alkyl phosphonate/methyl phosphonate internucleoside, which according to EP 2 742 135 may for example be tolerated in an otherwise DNA phosphorothioate the gap region.

Nucleobase

[0047] The term nucleobase includes the purine (e.g. adenine and guanine) and pyrimidine (e.g. uracil, thymine and cytosine) moiety present in nucleosides and nucleotides which form hydrogen bonds in nucleic acid hybridization. In the context of the present invention the term nucleobase also encompasses modified nucleobases which may differ from naturally occurring nucleobases, but are functional during nucleic acid hybridization. In this context "nucleobase"

refers to both naturally occurring nucleobases such as adenine, guanine, cytosine, thymidine, uracil, xanthine and hypoxanthine, as well as non-naturally occurring variants. Such variants are for example described in Hirao et al (2012) Accounts of Chemical Research vol 45 page 2055 and Bergstrom (2009) Current Protocols in Nucleic Acid Chemistry Suppl. 37 1.4.1.

[0048] In some embodiments the nucleobase moiety is modified by changing the purine or pyrimidine into a modified purine or pyrimidine, such as substituted purine or substituted pyrimidine, such as a nucleobase selected from isocytosine, pseudoisocytosine, 5-methyl cytosine, 5-thiozolo-cytosine, 5-propynyl-cytosine, 5-propynyl-uracil, 5-bromouracil 5-thiazolo-uracil, 2-thio-uracil, 2'thio-thymine, inosine, diaminopurine, 6-aminopurine, 2-aminopurine, 2,6-diaminopurine and 2-chloro-6-aminopurine.

[0049] The nucleobase moieties may be indicated by the letter code for each corresponding nucleobase, e.g. A, T, G, C or U, wherein each letter may optionally include modified nucleobases of equivalent function. For example, in the exemplified oligonucleotides, the nucleobase moieties are selected from A, T, G, C, and 5-methyl cytosine. Optionally, for LNA gapmers, 5-methyl cytosine LNA nucleosides may be used.

Modified Oligonucleotide

[0050] The term modified oligonucleotide or modified nucleic acid molecule describes an oligonucleotide or nucleic acid molecule comprising one or more sugar-modified nucleosides and/or modified internucleoside linkages. The term "chimeric" is a term that has been used in the literature to describe oligonucleotides or nucleic acid molecules with modified nucleosides, in particular gapmer oligonucleotides.

Complementarity

[0051] The term "complementarity" describes the capacity for Watson-Crick base-pairing of nucleosides/nucleotides. Watson-Crick base pairs are guanine (G)-cytosine (C) and adenine (A)-thymine (T)/uracil (U). It will be understood that oligonucleotides may comprise nucleosides with modified nucleobases, for example 5-methyl cytosine is often used in place of cytosine, and as such the term complementarity encompasses Watson Crick base-paring between non-modified and modified nucleobases (see for example Hirao et al (2012) Accounts of Chemical Research vol 45 page 2055 and Bergstrom (2009) Current Protocols in Nucleic Acid Chemistry Suppl. 37 1.4.1).

[0052] The term "% complementary" as used herein, refers to the proportion of nucleotides (in percent) of a contiguous nucleotide sequence in a nucleic acid molecule (e.g. oligonucleotide) which across the contiguous nucleotide sequence, are complementary to a reference sequence (e.g. a target sequence or sequence motif). The percentage of complementarity is thus calculated by counting the number of aligned nucleobases that are complementary (from Watson Crick base pair) between the two sequences (when aligned with the target sequence 5'-3' and the oligonucleotide sequence from 3'-5'), dividing that number by the total number of nucleotides in the oligonucleotide and multiplying by 100. In such a comparison, a nucleobase/nucleotide which does not align (form a base pair) is termed a mismatch. Insertions and deletions are not allowed in the

calculation of % complementarity of a contiguous nucleotide sequence. It will be understood that in determining complementarity, chemical modifications of the nucleobases are disregarded as long as the functional capacity of the nucleobase to form Watson Crick base pairing is retained (e.g. 5'-methyl cytosine is considered identical to a cytosine for the purpose of calculating % identity).

[0053] The term "fully complementary", refers to 100% complementarity.

[0054] The following is an example of an oligonucleotide that is fully complementary to the target nucleic acid.

[0055] The following is an example of an oligonucleotide motif (SEQ ID NO: 33) that is fully complementary to the target nucleic acid (SEQ ID NO: 22).

```
                                    (SEQ ID NO: 22)
              5' gagaaguucggaaugagua 3'

                                    (SEQ ID NO: 33)
              3'    cttcaagccttactc 5'
```

Identity

[0056] The term "Identity" as used herein, refers to the proportion of nucleotides (expressed in percent) of a contiguous nucleotide sequence in a nucleic acid molecule (e.g. oligonucleotide) which across the contiguous nucleotide sequence, are identical to a reference sequence (e.g. a sequence motif). The percentage of identity is thus calculated by counting the number of aligned nucleobases that are identical (a Match) between two sequences (in the contiguous nucleotide sequence of the compound of the invention and in the reference sequence), dividing that number by the total number of nucleotides in the oligonucleotide and multiplying by 100. Therefore, Percentage of Identity= (Matches×100)/Length of aligned region (e.g. the contiguous nucleotide sequence). Insertions and deletions are not allowed in the calculation the percentage of identity of a contiguous nucleotide sequence. It will be understood that in determining identity, chemical modifications of the nucleobases are disregarded as long as the functional capacity of the nucleobase to form Watson Crick base pairing is retained (e.g. 5-methyl cytosine is considered identical to a cytosine for the purpose of calculating % identity).

Hybridization

[0057] The term "hybridizing" or "hybridizes" as used herein is to be understood as two nucleic acid strands (e.g. an antisense oligonucleotide or siRNA guide strand and a target nucleic acid) forming hydrogen bonds between base pairs on opposite strands thereby forming a duplex. The affinity of the binding between two nucleic acid strands is the strength of the hybridization. It is often described in terms of the melting temperature (T_m) defined as the temperature at which half of the oligonucleotides are duplexed with the target nucleic acid. At physiological conditions T_m is not strictly proportional to the affinity (Mergny and Lacroix, 2003, *Oligonucleotides* 13:515-537). The standard state Gibbs free energy $\Delta G°$ is a more accurate representation of binding affinity and is related to the dissociation constant (K_d) of the reaction by $\Delta G°=-RT\ln(K_d)$, where R is the gas constant and T is the absolute temperature. Therefore, a very low $\Delta G°$ of the reaction between an oligonucleotide and the target nucleic acid reflects a strong

hybridization between the oligonucleotide and target nucleic acid. $\Delta G°$ is the energy associated with a reaction where aqueous concentrations are 1M, the pH is 7, and the temperature is 37° C. The hybridization of oligonucleotides to a target nucleic acid is a spontaneous reaction and for spontaneous reactions $\Delta G°$ is less than zero. $\Delta G°$ can be measured experimentally, for example, by use of the iso-thermal titration calorimetry (ITC) method as described in Hansen et al., 1965, Chem. Comm. 36-38 and Holdgate et al., 2005, Drug Discov Today. The skilled person will know that commercial equipment is available for $\Delta G°$ measurements. $\Delta G°$ can also be estimated numerically by using the nearest neighbor model as described by SantaLucia, 1998, *Proc Natl Acad Sci USA*. 95: 1460-1465 using appropriately derived thermodynamic parameters described by Sugimoto et al., 1995, *Biochemistry* 34:11211-11216 and McTigue et al., 2004, *Biochemistry* 43:5388-5405. In order to have the possibility of modulating its intended nucleic acid target by hybridization, oligonucleotides of the present invention hybridize to a target nucleic acid with estimated $\Delta G°$ values below −10 kcal for oligonucleotides that are 10-30 nucleotides in length. In some embodiments the degree or strength of hybridization is measured by the standard state Gibbs free energy $\Delta G°$. The oligonucleotides may hybridize to a target nucleic acid with estimated $\Delta G°$ values below the range of −10 kcal, such as below −15 kcal, such as below −20 kcal and such as below −25 kcal for oligonucleotides that are 8-30 nucleotides in length. In some embodiments the oligonucleotides hybridize to a target nucleic acid with an estimated $\Delta G°$ value of −10 to −60 kcal, such as −12 to −40, such as from −15 to −30 kcal or −16 to −27 kcal such as −18 to −25 kcal.

Target Nucleic Acid

[0058] According to the present invention, the target nucleic acid is a nucleic acid which encodes mammalian FUBP1 and may for example be a gene, a RNA, a mRNA, and pre-mRNA, a mature mRNA or a cDNA sequence. The target may therefore be referred to as a FUBP1 target nucleic acid.

[0059] The therapeutic nucleic acid molecules of the invention may for example target exon regions of a mammalian FUBP1 (in particular siRNA and shRNA, but also antisense oligonucleotides), or may for example target any intron region in the FUBP1 pre-mRNA (in particular antisense oligonucleotides). Table 1 lists predicted exon and intron regions of SEQ ID NO: 1.

TABLE 1

Exon and intron regions in the human FUBP1 pre-mRNA.

Exonic regions in the human FUBP1 premRNA (SEQ ID NO 1)			Intronic regions in the human FUBP1 premRNA (SEQ ID NO 1)		
ID	start	end	ID	start	end
E1	19	226	I1	227	9095
E2	9096	9186	I2	9187	10907
E3	10908	10946	I3	10947	11444
E4	11445	11484	I4	11485	12009
E5	12010	12062	I5	12063	12155
E6	12156	12227	I6	12228	12359
E7	12360	12417	I7	12418	13879
E8	13880	14042	I8	14043	14142

TABLE 1-continued

Exon and intron regions in the human FUBP1 pre-mRNA.

Exonic regions in the human FUBP1 premRNA (SEQ ID NO 1)			Intronic regions in the human FUBP1 premRNA (SEQ ID NO 1)		
ID	start	end	ID	start	end
E9	14143	14241	I9	14242	14363
E10	14364	14465	I10	14466	14754
E11	14755	14857	I11	14858	14948
E12	14949	15049	I12	15050	15395
E13	15396	15537	I13	15538	16180
E14	16181	16341	I14	16342	18615
E15	18616	18767	I15	18768	18847
E16	18848	18927	I16	18928	22410
E17	22411	22539	I17	22540	23781
E18	23782	23856	I18	23857	29810
E19	29811	29956	I19	29957	30196
E20	30197	30706			

[0060] Suitably, the target nucleic acid encodes a FUBP1 protein, in particular mammalian FUBP1, such as human FUBP1(See for example Tables 2 and 3) which provides the genomic sequence, the mature mRNA and pre-mRNA sequences for human, monkey and mouse FUBP1).

[0061] In some embodiments, the target nucleic acid is selected from the group consisting of SEQ ID NO: 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and/or 20, or naturally occurring variants thereof (e.g. sequences encoding a mammalian FUBP1).

[0062] In some embodiments, the target nucleic acid is selected from the group consisting of SEQ ID NO: 1, 5 and/or 9, or naturally occurring variants thereof (e.g. sequences encoding a mammalian FUBP1).

[0063] In some embodiments, the target nucleic acid is selected from the group consisting of SEQ ID NO: 1 and 5, or naturally occurring variants thereof (e.g. sequences encoding a mammalian FUBP1).

[0064] In some embodiments, the target nucleic acid is selected from the group consisting of SEQ ID NO: 1, to 8, or naturally occurring variants thereof (e.g. sequences encoding a mammalian FUBP1).

TABLE 2

Genome and assembly information for FUBP1 across species.

Species	Chr.	Strand	Genomic coordinates		ensembl	
			Start	End	Assembly	gene_id
Human	1	Rv	77944055	77979110	GRCh38. p10	ENSG 00000162613
Cyno monkey	1	Fwd	149243675	149283374	Macaca_ fascicu- laris_ 5.0	ENSMFAG 00000031825
Mouse	3	Fwd	152210422	152236826	GRCm38. p5	ENSMUSG 00000028034

Fwd = forward strand.
Rv = reverse strand.
The genome coordinates provide the pre-mRNA sequence (genomic sequence).

[0065] If employing the nucleic acid molecule of the invention in research or diagnostics the target nucleic acid may be a cDNA or a synthetic nucleic acid derived from DNA or RNA.

[0066] For in vivo or in vitro application, the therapeutic nucleic acid molecule of the invention is typically capable of inhibiting the expression of the FUBP1 target nucleic acid in a cell which is expressing the FUBP1 target nucleic acid. The contiguous sequence of nucleobases of the nucleic acid molecule of the invention is typically complementary to a conserved region of the FUBP1 target nucleic acid, as measured across the length of the oligonucleotide, optionally with the exception of one or two mismatches, and optionally excluding nucleotide based linker regions which may link the oligonucleotide to an optional functional group such as a conjugate, or other non-complementary terminal nucleotides. Further information on exemplary target nucleic acids is provided in table 3.

TABLE 3

Sequence details for FUBP1 across species.

Species	RNA type	Length (nt)	SEQ ID NO
Human	Pre-mRNA	305056	1
Human	Mature mRNA	696	2
Human	Mature mRNA	1968	3
Human	Mature mRNA	1935	3
Cyno monkey	Pre-mRNA	39750	5
Cyno monkey	Mature mRNA	1968	6
Cyno monkey	Mature mRNA	6825	7
Cyno monkey	Mature mRNA	1959	8
Mouse	Pre-mRNA	26405	9
Mouse	Mature mRNA	4525	10
Mouse	Mature mRNA	800	11
Mouse	Mature mRNA	2526	12
Mouse	Mature mRNA	809	13
Mouse	Mature mRNA	1040	14
Mouse	Mature mRNA	796	15
Mouse	Mature mRNA	585	16
Mouse	Mature mRNA	2374	17
Mouse	Mature mRNA	3163	18
Mouse	Mature mRNA	6523	19
Mouse	Mature mRNA	2552	20

[0067] Note SEQ ID NO 5 comprises regions of multiple NNNNs, where the sequencing has been unable to accurately refine the sequence, and a degenerate sequence is therefore included. For the avoidance of doubt the compounds of the invention are complementary to the actual target sequence and are not therefore degenerate compounds.

Target Sequence

[0068] The term "target sequence" as used herein refers to a sequence of nucleotides present in the target nucleic acid which comprises the nucleobase sequence which is complementary to the oligonucleotide or nucleic acid molecule of the invention. In some embodiments, the target sequence consists of a region on the target nucleic acid with a nucleobase sequence that is complementary to the contiguous nucleotide sequence of the oligonucleotide of the invention (i.e. a sub-sequence). This region of the target nucleic acid may interchangeably be referred to as the target nucleotide sequence, target sequence or target region. In some embodiments the target sequence is longer than the complementary sequence of a single nucleic acid molecule, and may, for example represent a preferred region of the target nucleic acid which may be targeted by several nucleic acid molecules of the invention.

[0069] In some embodiments the target sequence is a sequence selected from the group consisting of a human FUBP1 mRNA exon, such as a FUBP1 human mRNA exon selected from the group consisting of e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, e11, e12, 13, e14, e15, e16, e17, e18, e19 and e20 (see for example table 1 above).

[0070] In one embodiment the target sequence is a sequence selected from the group consisting of one or more of human FUBP1 mRNA exons selected from the group consisting of exon 9, 10, 12 and 20.

[0071] In some embodiments the target sequence is a sequence selected from the group consisting of a human FUBP1mRNA intron, such as a FUBP1 human mRNA intron selected from the group consisting of i1, i2, i3, 14, 15, 16, i7, 19, i10, i11, i12, 13, i14, 115, i16, i17, i18 and i19 (see for example table 1 above).

[0072] The nucleic acid molecule of the invention comprises a contiguous nucleotide sequence which is complementary to or hybridizes to a region on the target nucleic acid, such as a target sequence described herein.

[0073] The target nucleic acid sequence to which the therapeutic nucleic acid molecule is complementary or hybridizes to generally comprises a stretch of contiguous nucleobases of at least 10 nucleotides. The contiguous nucleotide sequence is between 12 to 70 nucleotides, such as 12-50, such as 13 to 30, such as 14 to 25, such as 15 to 20, such as 16 to 18 contiguous nucleotides.

Target Cell

[0074] The term a "target cell" as used herein refers to a cell which is expressing the target nucleic acid. For the therapeutic use of the present invention it is advantageous if the target cell is infected with HBV. In some embodiments the target cell may be in vivo or in vitro. In some embodiments the target cell is a mammalian cell such as a rodent cell, such as a mouse cell or a rat cell, or a primate cell such as a monkey cell or a human cell.

[0075] In preferred embodiments the target cell expresses FUBP1 mRNA, such as the FUBP1pre-mRNA or FUBP1 mature mRNA. The poly A tail of FUBP1 mRNA is typically disregarded for nucleic acid molecule targeting.

Naturally Occurring Variant

[0076] The term "naturally occurring variant" refers to variants of theFUBP1 gene or transcripts which originate from the same genetic loci as the target nucleic acid, but may differ for example, by virtue of degeneracy of the genetic code causing a multiplicity of codons encoding the same amino acid, or due to alternative splicing of pre-mRNA, or the presence of polymorphisms, such as single nucleotide polymorphisms, and allelic variants. Based on the presence of the sufficient complementary sequence to the oligonucleotide, the oligonucleotide of the invention may therefore target the target nucleic acid and naturally occurring variants thereof.

[0077] In some embodiments, the naturally occurring variants have at least 95% such as at least 98% or at least 99% homology to a mammalian FUBP1 target nucleic acid, such as a target nucleic acid selected form the group consisting of SEQ ID NO: 1, 5 or 9. In some embodiments the naturally occurring variants have at least 99% homology to the human FUBP1 target nucleic acid of SEQ ID NO: 1. In some embodiments the naturally occurring variants are known polymorphisms.

Modulation of Expression

[0078] The term "modulation of expression" as used herein is to be understood as an overall term for a nucleic acid molecules ability to alter the amount of FUBP1 when compared to the amount of FUBP1 before administration of the nucleic acid molecule. Alternatively, modulation of expression may be determined by reference to a control experiment. It is generally understood that the control is an individual or target cell treated with a saline composition or an individual or target cell treated with a non-targeting or nucleic acid molecule (mock). It may however also be an individual treated with the standard of care.

[0079] One type of modulation is the ability of a nucleic acid molecule to inhibit, down-regulate, reduce, remove, stop, prevent, lessen, lower, avoid or terminate expression of FUBP1, e.g. by degradation of mRNA or blockage of transcription.

High Affinity Modified Nucleosides

[0080] A high affinity modified nucleoside is a modified nucleotide which, when incorporated into the oligonucleotide enhances the affinity of the oligonucleotide for its complementary target, for example as measured by the melting temperature (T^m). A high affinity modified nucleoside of the present invention preferably result in an increase in melting temperature between +0.5 to +12° C. , more preferably between +1.5 to +10° C. and most preferably between +3 to +8° C. per modified nucleoside. Numerous high affinity modified nucleosides are known in the art and include for example, many 2' sugar modified nucleosides, such as 2' substituted nucleosides like Ome and MOE as well as 2' to 4' bridged nucleic acids such as locked nucleic acids (LNA) (see e.g. Freier & Altmann; Nucl. Acid Res., 1997, 25, 4429-4443 and Uhlmann; Curr. Opinion in Drug Development, 2000, 3(2), 293-213).

Sugar Modifications

[0081] The nucleic acid molecule of the invention may comprise one or more nucleosides which have a modified sugar moiety, i.e. a modification of the sugar moiety when compared to the ribose sugar moiety found in DNA and RNA.

[0082] Numerous nucleosides with modification of the ribose sugar moiety have been made, primarily with the aim of improving certain properties of nucleic acid molecules, such as affinity and/or nuclease resistance.

[0083] Such modifications include those where the ribose ring structure is modified, e.g. by replacement with a hexose ring (HNA), or a bicyclic ring, which typically have a biradicle bridge between the C2 and C4 carbons on the ribose ring (LNA), or an unlinked ribose ring which typically lacks a bond between the C2 and C3 carbons (e.g. UNA). Other sugar modified nucleosides include, for example, bicyclohexose nucleic acids (WO2011/017521) or tricyclic nucleic acids (WO2013/154798). Modified nucleosides also include nucleosides where the sugar moiety is replaced with a non-sugar moiety, for example in the case of peptide nucleic acids (PNA), or morpholino nucleic acids.

[0084] Sugar modifications also include modifications made via altering the substituent groups on the ribose ring to groups other than hydrogen, or the —OH groups naturally found in RNA or DNA nucleosides. Substituents may, for example be introduced at the 2', 3', 4' or 5' positions.

2' Sugar Modified Nucleosides

[0085] A 2' sugar modified nucleoside is a nucleoside which has a substituent other than H or —OH at the 2' position (2' substituted nucleoside) or comprises a 2' linked biradicle capable of forming a bridge between the 2' carbon and a second carbon in the ribose ring, such as LNA (2'-4' biradicle bridged) nucleosides.

[0086] Indeed, much focus has been spent on developing 2' substituted nucleosides, and numerous 2' substituted nucleosides have been found to have beneficial properties when incorporated into oligonucleotides. For example, the 2' modified sugar may provide enhanced binding affinity and/ or increased nuclease resistance to the oligonucleotide. Examples of 2' substituted modified nucleosides are 2'-O-alkyl-RNA, 2'-O-methyl-RNA, 2'-alkoxy-RNA, 2'-O-methoxyethyl-RNA (MOE), 2'-amino-DNA, 2'-Fluoro-RNA, and 2'-F-ANA nucleoside. For further examples, please see e.g. Freier & Altmann; Nucl. Acid Res., 1997, 25, 4429-4443 and Uhlmann; Curr. Opinion in Drug Development, 2000, 3(2), 293-213, and Deleavey and Damha, Chemistry and Biology 2012, 19, 937. Below are illustrations of some 2' substituted modified nucleosides.

2'-O-Me

2'F-RNA

2'F-ANA

2'-O-MOE

2'-O-Allyl

2'-O-Ethylamine
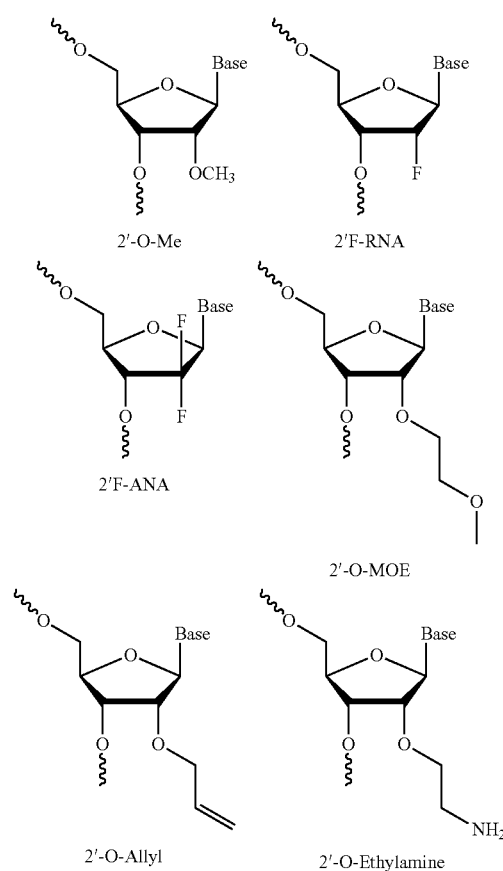
[0087] In relation to the present invention 2' substituted does not include 2' bridged molecules like LNA.

Locked Nucleic Acid Nucleosides (LNA).

[0088] A "LNA nucleoside" is a 2'-modified nucleoside which comprises a biradical linking the C2' and C4' of the

ribose sugar ring of said nucleoside (also referred to as a "2'-4' bridge"), which restricts or locks the conformation of the ribose ring. These nucleosides are also termed bridged nucleic acid or bicyclic nucleic acid (BNA) in the literature. The locking of the conformation of the ribose is associated with an enhanced affinity of hybridization (duplex stabilization) when the LNA is incorporated into an oligonucleotide for a complementary RNA or DNA molecule. This can be routinely determined by measuring the melting temperature of the oligonucleotide/complement duplex.

[0089] Non limiting, exemplary LNA nucleosides are disclosed in WO 99/014226, WO 00/66604, WO 98/039352, WO 2004/046160, WO 00/047599, WO 2007/134181, WO 2010/077578, WO 2010/036698, WO 2007/090071, WO 2009/006478, WO 2011/156202, WO 2008/154401, WO 2009/067647, WO 2008/150729, Morita et al., Bioorganic & Med. Chem. Lett. 12, 73-76, Seth et al. J. Org. Chem. 2010, Vol 75(5) pp. 1569-81, and Mitsuoka et al., Nucleic Acids Research 2009, 37(4), 1225-1238, and Wan and Seth, J. Medical Chemistry 2016, 59, 9645-9667.

[0090] Further non limiting, exemplary LNA nucleosides are disclosed in Scheme 1.

Scheme 1

β-D-oxy LNA β-D-amino LNA β-D-thio LNA

β-D-amino substituted LNA α-L-oxy LNA α-L-amino LNA

α-L-thio LNA 6'-methyl-β-D-oxy LNA 6'-dimethyl-β-D-oxy LNA

5'-methyl-β-D-oxy LNA 5'-methyl-6'-dimethyl-β-D-oxy LNA

-continued

carbocyclic(vinyl)-β-D-oxy LNA carbocyclic(vinyl)-α-L-oxy LNA

6'-methyl-β-D-thio LNA β-D-amino substituted LNA

ENA COC (S)-cET
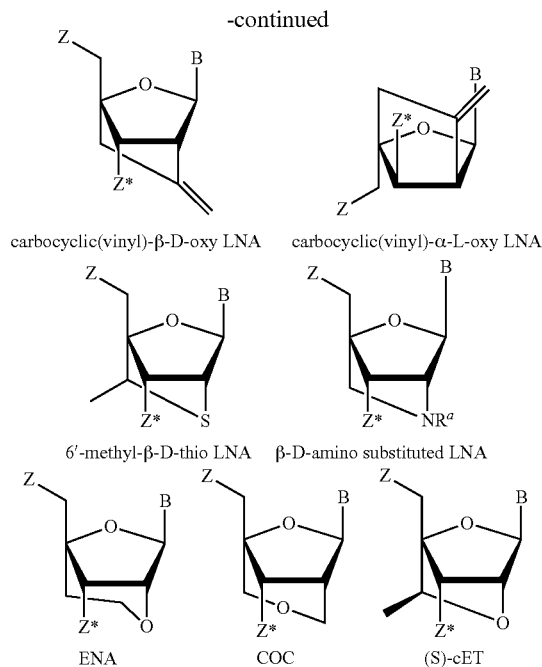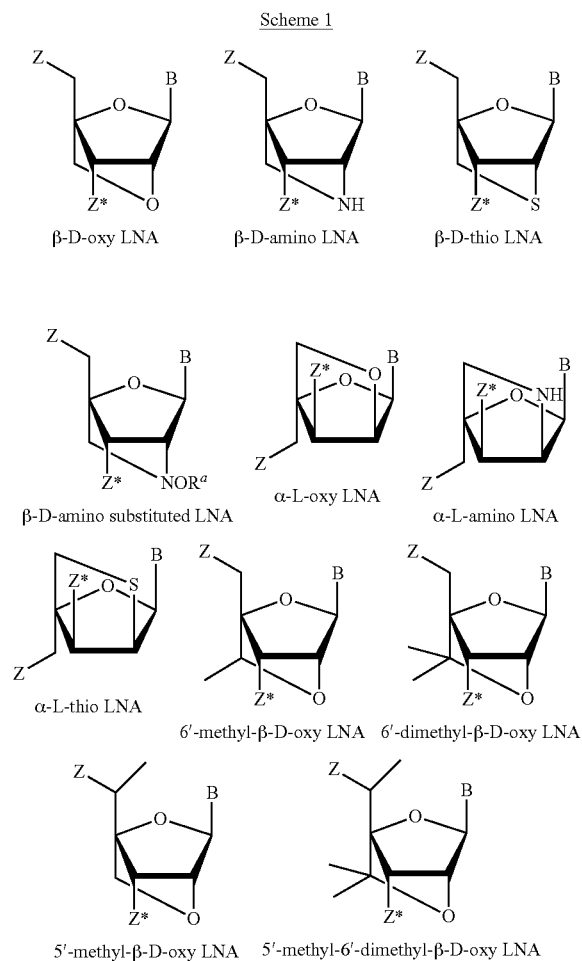
[0091] Particular LNA nucleosides are beta-D-oxy-LNA, 6'-methyl-beta-D-oxy LNA such as (S)-6'-methyl-beta-D-oxy-LNA (ScET) and ENA. A particularly advantageous LNA is beta-D-oxy-LNA.

Nuclease Mediated Degradation

[0092] Nuclease mediated degradation refers to nucleic acid molecule capable of mediating degradation of a complementary nucleotide sequence when forming a duplex with such a sequence.

[0093] In some embodiments, the nucleic acid molecule may function via nuclease mediated degradation of the target nucleic acid. In particular, antisense oligonucleotides are capable of recruiting an endoribonuclease (RNase) which recognizes RNA/DNA hybridization and effects cleavage of the RNA nucleic acid. RNase H has proven to be an advantageous endoribonuclease. Examples of antisense oligonucleotide designs which operate via nuclease mediated mechanisms are oligonucleotides which typically comprise a region of at least 5 or 6 consecutive DNA nucleosides and are flanked on one side or both sides by affinity enhancing nucleosides, for example gapmers, headmers and tailmers.

RNase H Activity and Recruitment

[0094] The RNase H activity of an antisense oligonucleotide refers to its ability to recruit RNase H when in a duplex with a complementary RNA molecule. WO01/23613 provides in vitro methods for determining RNaseH activity, which may be used to determine the ability to recruit RNaseH. Typically an oligonucleotide is deemed capable of recruiting RNase H if it, when provided with a complementary target nucleic acid sequence, has an initial rate, as measured in pmol/l/min, of at least 5%, such as at least 10% or more than 20% of the of the initial rate determined when using a oligonucleotide having the same base sequence as the modified oligonucleotide being tested, but containing

only DNA monomers with phosphorothioate linkages between all monomers in the oligonucleotide, and using the methodology provided by Example 91-95 of WO01/23613 (hereby incorporated by reference). For use in determining RHase H activity, recombinant human RNase H1 is available from Lubio Science GmbH, Lucerne, Switzerland

Gapmer

[0095] The nucleic acid molecule of the invention, or contiguous nucleotide sequence thereof may be a gapmer antisense oligonucleotide. The antisense gapmers are commonly used to inhibit a target nucleic acid via RNase H mediated degradation. In an embodiment of the invention the oligonucleotide of the invention is capable of recruiting RNase H.

[0096] A gapmer antisense oligonucleotide comprises at least three distinct structural regions a 5'-flank, a gap and a 3'-flank, F-G-F' in the '5 ->3' orientation. The "gap" region (G) comprises a stretch of contiguous DNA nucleotides which enable the oligonucleotide to recruit RNase H. The gap region is flanked by a 5' flanking region (F) comprising one or more sugar modified nucleosides, advantageously high affinity sugar modified nucleosides, and by a 3' flanking region (F') comprising one or more sugar modified nucleosides, advantageously high affinity sugar modified nucleosides. The one or more sugar modified nucleosides in region F and F' enhance the affinity of the oligonucleotide for the target nucleic acid (i.e. are affinity enhancing sugar modified nucleosides). In some embodiments, the one or more sugar modified nucleosides in region F and F' are 2' sugar modified nucleosides, such as high affinity 2' sugar modifications, such as independently selected from LNA and 2'-MOE.

[0097] In a gapmer design, the 5' and 3' most nucleosides of the gap region are DNA nucleosides, and are positioned adjacent to a sugar modified nucleoside of the 5' (F) or 3' (F') region respectively. The flanks may further be defined by having at least one sugar modified nucleoside at the end most distant from the gap region, i.e. at the 5' end of the 5' flank and at the 3' end of the 3' flank.

[0098] Regions F-G-F' form a contiguous nucleotide sequence. Antisense oligonucleotides of the invention, or the contiguous nucleotide sequence thereof, may comprise a gapmer region of formula F-G-F'.

[0099] The overall length of the gapmer design F-G-F' may be, for example 12 to 32 nucleosides, such as 13 to 24, such as 14 to 22 nucleosides, Such as from 14 to 17, such as 16 to 18 nucleosides. By way of example, the gapmer oligonucleotide of the present invention can be represented by the following formulae:

$$F_{1\text{-}8}\text{-}G_{5\text{-}16}\text{-}F'_{1\text{-}8}, \text{ such as}$$

$$F_{1\text{-}8}\text{-}G_{7\text{-}16}\text{-}F'_{2\text{-}8}$$
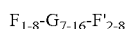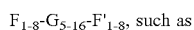
with the proviso that the overall length of the gapmer regions F-G-F' is at least 12, such as at least 14 nucleotides in length.

[0100] In an aspect of the invention the antisense oligonucleotide or contiguous nucleotide sequence thereof consists of or comprises a gapmer of formula 5'-F-G-F'-3', where region F and F' independently comprise or consist of 1-8 nucleosides, of which 1-4 are 2' sugar modified and defines the 5' and 3' end of the F and F' region, and G is a region between 6 and 16 nucleosides which are capable of recruiting RNaseH.

[0101] In some embodiments the gap region G may consist of 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16 contiguous phosphorothioate linked DNA nucleosides. In some embodiments, all internucleoside linkages in the gap are phosphorothioate linkages.

[0102] In some embodiments, region F and F' independently consists of or comprises a contiguous sequence of sugar modified nucleosides. In some embodiments, the sugar modified nucleosides of region F may be independently selected from 2'-O-alkyl-RNA units, 2'-O-methyl-RNA, 2'-amino-DNA units, 2'-fluoro-DNA units, 2'-alkoxy-RNA, MOE units, LNA units, arabino nucleic acid (ANA) units and 2'-fluoro-ANA units.

[0103] In some embodiments, all the nucleosides of region F or F', or F and F' are LNA nucleosides, such as independently selected from beta-D-oxy LNA, ENA or ScET nucleosides. In some embodiments region F consists of 1-5, such as 2-4, such as 3-4 such as 1, 2, 3, 4 or 5 contiguous LNA nucleosides. In some embodiments, all the nucleosides of region F and F' are beta-D-oxy LNA nucleosides.

[0104] In some embodiments, all the nucleosides of region F or F', or F and F' are 2' substituted nucleosides, such as OMe or MOE nucleosides. In some embodiments region F consists of 1, 2, 3, 4, 5, 6, 7, or 8 contiguous OMe or MOE nucleosides. In some embodiments only one of the flanking regions can consist of 2' substituted nucleosides, such as OMe or MOE nucleosides. In some embodiments it is the 5' (F) flanking region that consists 2' substituted nucleosides, such as OMe or MOE nucleosides whereas the 3' (F') flanking region comprises at least one LNA nucleoside, such as beta-D-oxy LNA nucleosides or cET nucleosides. In some embodiments it is the 3' (F') flanking region that consists 2' substituted nucleosides, such as OMe or MOE nucleosides whereas the 5' (F) flanking region comprises at least one LNA nucleoside, such as beta-D-oxy LNA nucleosides or cET nucleosides.

[0105] In some embodiments, all the modified nucleosides of region F and F' are LNA nucleosides, such as independently selected from beta-D-oxy LNA, ENA or ScET nucleosides, wherein region F or F', or F and F' may optionally comprise DNA nucleosides (an alternating flank, see definition of these for more details). In some embodiments, all the modified nucleosides of region F and F' are beta-D-oxy LNA nucleosides, wherein region F or F', or F and F' may optionally comprise DNA nucleosides (an alternating flank, see definition of these for more details).

[0106] Further gapmer designs are disclosed in WO2004/046160, WO2007/146511 and WO2008/113832, hereby incorporated by reference.

Conjugate

[0107] The term conjugate as used herein refers to an oligonucleotide which is covalently linked to a non-nucleotide moiety (conjugate moiety or region C or third region).

[0108] Conjugation of the therapeutic nucleic acid molecule of the invention to one or more non-nucleotide moieties may improve the pharmacology of the nucleic acid molecule, e.g. by affecting the activity, cellular distribution, cellular uptake or stability of the nucleic acid molecule. In some embodiments the conjugate moiety, modify or enhance the pharmacokinetic properties of the nucleic acid molecule by improving cellular distribution, bioavailability, metabolism, excretion, permeability, and/or cellular uptake of the oligonucleotide. In particular, the conjugate may target the

nucleic acid molecule to a specific organ, tissue or cell type and thereby enhance the effectiveness of the oligonucleotide in that organ, tissue or cell type. At the same time the conjugate may serve to reduce activity of the nucleic acid molecule in non-target cell types, tissues or organs, e.g. off target activity or activity in non-target cell types, tissues or organs. For siRNA nucleic acid molecules the conjugate moiety is most commonly covalently linked to the passenger strand of the siRNA, and for shRNA molecules the conjugate moiety would most commonly be linked to the end of the molecule which is furthest away from the contiguous nucleotide sequence of the shRNA. For antisense oligonucleotides the conjugate moiety can be covalently linked to any of the terminal ends, advantageously using a biocleavable linker such as a 2 to 5 phosphodiester linked DNA nucleosides.

[0109] WO 93/07883 and WO2013/033230 provides suitable conjugate moieties, which are hereby incorporated by reference. Further suitable conjugate moieties are those capable of binding to the asialoglycoprotein receptor (ASGPR). In particular tri-valent N-acetylgalactosamine conjugate moieties are suitable for binding to the ASGPR, see for example US 2009/02398, WO 2014/076196, WO 2014/207232 and WO 2014/179620 (hereby incorporated by reference).

[0110] Such conjugates serve to enhance uptake of the therapeutic nucleic acid molecule to the liver while reducing its presence in the kidney, thereby increasing the liver/kidney ratio of a conjugated nucleic acid molecule compared to the unconjugated version of the same nucleic acid molecule.

[0111] In an embodiment, the non-nucleotide moiety (conjugate moiety) is selected from the group consisting of carbohydrates, cell surface receptor ligands, drug substances, hormones, lipophilic substances, polymers, proteins, peptides, toxins (e.g. bacterial toxins), vitamins, viral proteins (e.g. capsids) or combinations thereof.

Conjugate Linkers

[0112] A linkage or linker is a connection between two atoms that links one chemical group or segment of interest to another chemical group or segment of interest via one or more covalent bonds. Conjugate moieties can be attached to the oligonucleotide directly or through a linking moiety (e.g. linker or tether). Linkers serve to covalently connect one region, e.g. a conjugate moiety to another region, e.g. an oligonucleotide (e.g. the termini of region A or C).

[0113] In some embodiments of the invention the conjugate or therapeutic nucleic acid molecule conjugate of the invention may optionally, comprise a linker region which is positioned between the oligonucleotide and the conjugate moiety. In some embodiments, the linker between the conjugate and oligonucleotide is physiologically labile linker, interchangeably termed biocleavable linker. The linker and the oligonucleotide is often attached via a phosphodiester linkage.

[0114] Biocleavable linkers (Region B) comprising or consisting of a physiologically labile bond that is cleavable under conditions normally encountered or analogous to those encountered within a mammalian body. Conditions under which physiologically labile linkers undergo chemical transformation (e.g., cleavage) include chemical conditions such as pH, temperature, oxidative or reductive conditions or agents, and salt concentration found in or analogous to

those encountered in mammalian cells. Mammalian intracellular conditions also include the presence of enzymatic activity normally present in a mammalian cell such as from proteolytic enzymes or hydrolytic enzymes or nucleases. In one embodiment the biocleavable linker is susceptible to S1 nuclease cleavage. In some embodiments the physiologically labile linker (biocleavable) comprises between 1 and 10 linked nucleosides, such as 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 linked nucleosides, such as between 2 and 6 linked nucleosides, such as between 2 and 5 linked nucleosides, such as between 2 and 4 linked nucleosides, where at least two consecutive linkages are biocleavable, such as phosphodiester linkages, such as at least 3 or 4 or 5 consecutive phosphodiester linkages. Preferably the nucleosides are DNA or RNA.

[0115] In one embodiment the linker between the oligonucleotide and the conjugate moiety is a physiologically labile linker composed of 2 to 5 consecutive phosphodiester linked nucleosides comprising at least two consecutive phosphodiester linkages at the 5' or 3' terminal of the contiguous nucleotide sequence of the antisense oligonucleotide or siRNA guide strand.

[0116] In another embodiment the linker between the oligonucleotide and the conjugate moiety is a physiologically labile linker composed of 2 to 5 consecutive phosphodiester linked nucleosides comprising at least two consecutive phosphodiester linkages at the 5' or 3' terminal of the passenger strand of the siRNA,

[0117] In another embodiment the linker between the oligonucleotide and the conjugate moiety is a physiologically labile linker composed of 2 to 5 consecutive phosphodiester linked nucleosides comprising at least two consecutive phosphodiester linkages at the terminal end of the shRNA molecule furthest away from the contiguous nucleotide sequence of the shRNA.

[0118] In some embodiments the physiologically labile linker comprises or consist of a DNA dinucleotide with a sequence selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, or GG, where there is a phosphodiester linkage between the two DNA nucleosides and at least one further phosphodiester at the 5' or 3' end of the dinucleotide linking either the oligonucleotide of the nucleic acid molecule to the dinucleotide or the conjugate moiety to the dinucleotide. In some embodiments the physiologically labile linker comprises or consist of a DNA trinucleotide of sequence AAA, AAT, AAC, AAG, ATA, ATT, ATC, ATG, ACA, ACT, ACC, ACG, AGA, AGT, AGC, AGG, TAA, TAT, TAC, TAG, TTA, TTT, TTC, TAG, TCA, TCT, TCC, TCG, TGA, TGT, TGC, TGG, CAA, CAT, CAC, CAG, CTA,

[0119] CTG, CTC, CTT, CCA, CCT, CCC, CCG, CGA, CGT, CGC, CGG, GAA, GAT, GAC, CAG, GTA, GTT, GTC, GTG, GCA, GCT, GCC, GCG, GGA, GGT, GGC, or GGG, where there is phosphodiester linkages between the DNA nucleosides and potentially a further phosphodiester at the 5' or 3' end of the trinucleotide. Phosphodiester containing biocleavable linkers are described in more detail in WO 2014/076195 (hereby incorporated by reference). In a conjugate compound with a biocleavable linker at least about 50% of the conjugate moiety is cleaved from the oligonucleotide, such as at least about 60% cleaved, such as at least about 70% cleaved, such as at least about 80% cleaved, such as at least about 85% cleaved, such as at least about 90%

cleaved, such as at least about 95% of the conjugate moiety is cleaved from the oligonucleotide cleaved when compared against a standard.

[0120] Conjugates may also be linked to the oligonucleotide via non-biocleavable linkers, or in some embodiments the conjugate may comprise a non-cleavable linker which is covalently attached to the biocleavable linker. Linkers that are not necessarily biocleavable primarily serve to covalently connect a conjugate moiety to the oligonucleotide or biocleavable linker, and potentially generate some distance between the conjugate moiety and the oligonucleotide. Some example linkers (region Y) include 8-amino-3,6-dioxaoctanoic acid (ADO), succinimidyl 4-(N-maleimidomethyl)cyclohexane-I-carboxylate (SMCC), 6-amino-hexanoic acid (AHEX or AHA), 6-aminohexyloxy, 4-aminobutyric acid, 4-aminocyclohexylcarboxylic acid, succinimidyl 4-(N-maleimidomethyl)cyclohexane-I-carboxy-(6-amido-caproate) (LCSMCC), succinimidyl m-maleimido-benzoylate (MBS), succinimidyl N-e-maleimido-caproylate (EMCS), succinimidyl 6-(beta-maleimido-propionamido) hexanoate (SMPH), succinimidyl N-(a-maleimido acetate) (AMAS), succinimidyl 4-(p-maleimidophenyl)butyrate (SMPB), beta-alanine (beta-ALA), phenylglycine (PHG), 4-aminocyclohexanoic acid (ACHC), beta-(cyclopropyl) alanine (beta-CYPR), amino dodecanoic acid (ADC), alylene diols, polyethylene glycols, amino acids, and the like. Non-cleavable linkers may also comprise a chain structure or an oligomer of repeating units such as ethylene glycol, amino acid units or amino alkyl groups. In some embodiments the linker (region Y) is an amino alkyl, such as a C_2-C_{36} amino alkyl group, including, for example C_6 to C_{12} amino alkyl groups. In some embodiments the linker (region Y) is a C_6 amino alkyl group (also termed a C6 linker). Conjugate linker groups may be routinely attached to an oligonucleotide via use of an amino modified oligonucleotide, and an activated ester group on the conjugate group. The linkage group between the amino alkyl and the oligonucleotide may for example be a phosphorothioate or a phosphodiester, or one of the other nucleoside linkage groups referred to herein. A conjugate compound of the present invention may be composed of the following regions C—B-A (Conjugate moiety-biocleavable linker-oligonucleotide/contiguous nucleotide sequence) or C—Y—B-A (conjugate moiety-non-cleavable linker-biocleavable linker-oligonucleotide/contiguous nucleotide sequence).

Treatment

[0121] The terms "treatment", "treating", "treats" or the like as used herein generally means obtaining a desired pharmacological and/or physiological effect. This effect is therapeutic in terms of partially or completely curing a disease and/or adverse effect attributed to the disease. The term "treatment" as used herein covers any treatment of a disease in a subject and includes: (a) inhibiting the disease, i.e. arresting its development like the inhibition of increase of HBsAg and/or HBeAg; or (b) ameliorating (i.e. relieving) the disease, i.e. causing regression of the disease, like the repression of HBsAg and/or HBeAg production. Thus, a compound that ameliorates and/or inhibits a HBV infection is a compound that treats a HBV invention. Preferably, the term "treatment" as used herein relates to medical intervention of an already manifested disorder, like the treatment of an already defined and manifested HBV infection.

Prevention

[0122] Herein the term "preventing", "prevention" or "prevents" relates to a prophylactic treatment, i.e. to a measure or procedure the purpose of which is to prevent, rather than to cure a disease. Prevention means that a desired pharmacological and/or physiological effect is obtained that is prophylactic in terms of completely or partially preventing a disease or symptom thereof. Accordingly, herein "preventing a HBV infection" includes preventing a HBV infection from occurring in a subject, and preventing the occurrence of symptoms of a HBV infection. In the present invention in particular the prevention of HBV infection in children from HBV infected mothers are contemplated. Also contemplated is the prevention of an acute HBV infection turning into a chronic HBV infection.

Patient

[0123] For the purposes of the present invention the "subject" (or "patient") may be a vertebrate. In context of the present invention, the term "subject" includes both humans and other animals, particularly mammals, and other organisms. Thus, the herein provided means and methods are applicable to both human therapy and veterinary applications. Accordingly, herein the subject may be an animal such as a mouse, rat, hamster, rabbit, guinea pig, ferret, cat, dog, chicken, sheep, bovine species, horse, camel, or primate. Preferably, the subject is a mammal. More preferably the subject is human.

HBV Infection

[0124] The term "hepatitis B virus infection" or "HBV infection" is commonly known in the art and refers to an infectious disease that is caused by the hepatitis B virus (HBV) and affects the liver. A HBV infection can be an acute or a chronic infection. Some infected persons have no symptoms during the initial infection and some develop a rapid onset of sickness with vomiting, yellowish skin, tiredness, dark urine and abdominal pain ("Hepatitis B Fact sheet N° 204". who.int. July 2014. Retrieved 4 Nov. 2014). Often these symptoms last a few weeks and can result in death. It may take 30 to 180 days for symptoms to begin. In those who get infected around the time of birth 90% develop a chronic hepatitis B infection while less than 10% of those infected after the age of five do ("Hepatitis B FAQs for the Public—Transmission", U.S. Centers for Disease Control and Prevention (CDC), retrieved 2011-11-29). Most of those with chronic disease have no symptoms; however, cirrhosis and liver cancer may eventually develop (Chang, 2007, Semin Fetal Neonatal Med, 12: 160-167). These complications result in the death of 15 to 25% of those with chronic disease ("Hepatitis B Fact sheet N° 204". who.int. July 2014, retrieved 4 Nov. 2014). Herein, the term "HBV infection" includes the acute and chronic hepatitis B infection. The term "HBV infection" also includes the asymptotic stage of the initial infection, the symptomatic stages, as well as the asymptotic chronic stage of the HBV infection.

cccDNA cccDNA (covalently closed circular DNA) is a special DNA structure that arises during the propagation of some DNA viruses (Polyomaviridae) in the cell nucleus. cccDNA is a double-stranded DNA that originates in a linear form that is ligated by means of DNA ligase to a covalently closed ring. In most cases, transcription of viral DNA can

occur from the circular form only. The cccDNA of viruses is also known as episomal DNA or occasionally as a minichromosome.

[0125] cccDNA is typical of Caulimoviridae and Hepadnaviridae, including the hepatitis B virus (HBV). The HBV genome forms a stable minichromosome, the covalently closed circular DNA (cccDNA), in the hepatocyte nucleus. The cccDNA is formed by conversion of capsid-associated relaxed circular DNA (rcDNA). HBV cccDNA formation involves a multi-step process that requires the cellular DNA repair machinery and relies on specific interactions with distinct cellular components that contribute to the completion of the positive strand DNA in rcDNA (Alweiss et al. 2017, Viruses, 9 (6): 156).

Pharmaceutically Acceptable Salts

[0126] The term "pharmaceutically acceptable salts" refers to those salts which retain the biological effectiveness and properties of the free bases or free acids, which are not biologically or otherwise undesirable. The salts are formed with inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid, particularly hydrochloric acid, and organic acids such as acetic acid, propionic acid, glycolic acid, pyruvic acid, oxalic acid, maleic acid, malonic acid, succinic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, p-toluenesulfonic acid, salicylic acid, N-acetylcystein. In addition, these salts may be prepared form addition of an inorganic base or an organic base to the free acid. Salts derived from an inorganic base include, but are not limited to, the sodium, potassium, lithium, ammonium, calcium, magnesium salts. Salts derived from organic bases include, but are not limited to salts of primary, secondary, and tertiary amines, substituted amines including naturally occurring substituted amines, cyclic amines and basic ion exchange resins, such as isopropylamine, trimethylamine, diethylamine, triethylamine, tripropylamine, ethanolamine, lysine, arginine, N-ethylpiperidine, piperidine, polyamine resins. The compound of formula (I) can also be present in the form of zwitterions. Particularly preferred pharmaceutically acceptable salts of compounds of formula (I) are the salts of hydrochloric acid, hydrobromic acid, sulfuric acid, phosphoric acid and methanesulfonic acid.

Inhibitor

[0127] The term "inhibitor" is known in the art and relates to a compound/substance or composition capable of fully or partially preventing or reducing the physiologic function (i.e. the activity) of (a) specific protein(s) (e.g. of FUBP1). In the context of the present invention, an "inhibitor" of FUBP1 is capable of preventing or reducing the activity/ function of FUBP1, respectively, by preventing or reducing the expression of the FUBP1 gene products. Thus, an inhibitor of FUBP1 may lead to a decreased expression level of FUBP1 (e.g. decreased level of FUBP1 mRNA, or of FUBP1protein) which is reflected in a decreased functionality (i.e. activity) of FUBP1, wherein said function comprises the poly-A polymerase function. An inhibitor of FUBP1, in the context of the present invention, accordingly, may also encompass transcriptional repressors of FUBP1 expression that are capable of reducing the level of FUBP1. Preferred inhibitors are nucleic acid molecules of the invention.

Compound

[0128] Herein, the term "compound" means any molecule capable of inhibition FUBP1 expression or activity. Particular compounds of the invention are nucleic acid molecules, such as RNAi molecules or antisense oligonucleotides according to the invention or any conjugate comprising such a nucleic acid molecule. For example, herein the compound may be a nucleic acid molecule targeting FUBP1, in particular an antisense oligonucleotide or a siRNA.

Composition

[0129] The term "composition" may also be used to describe a nucleic acid molecule compound or another FUBP1 inhibitor. A nucleic acid molecule composition has less than 20% impurities, preferably less than 15% or 10% impurities, more preferably less than 9, 8, 7 or 6% impurities, most preferably less than 5% impurities. The impurities are typically nucleic acid molecules which are one or two nucleotides shorter (n–1 or n–2) than the primary nucleic acid molecule component.

[0130] The present invention is further described by reference to the non-limiting figures and examples.

DETAILED DESCRIPTION OF THE INVENTION

[0131] Overexpression of and mutations in FUBP1 has been known to be associated with cancers for many years. In particular, strong overexpression of FUBP1 in human hepatocellular carcinoma

[0132] (HCC) supports tumour growth and correlates with poor patient prognosis. HBV cccDNA in infected hepatocytes is responsible for persistent chronic infection and reactivation, being the template for all viral subgenomic transcripts and pre-genomic RNA (pgRNA) to ensure both newly synthesized viral progeny and cccDNA pool replenishment via intracellular nucleocapsid recycling. In the context of the present invention it was for the first time shown that FUBP1 is associated with cccDNA stability. This knowledge allows for the opportunity to destabilize cccDNA in HBV infected subjects which in turn opens the opportunity for a complete cure of chronically infected HBV patients. The role of FUBP1 in HCC and cccDNA stability is expected to be different and independent of each other.

[0133] One aspect of the present invention is a FUBP1 inhibitor for use in the treatment and/or prevention of Hepatitis B virus (HBV) infection, in particular a chronic HBV infection.

[0134] An embodiment of the invention is a FUBP1 inhibitor which is capable of reducing cccDNA and pgRNA in an infected cell, such as an HBV infected cell.

[0135] In a further embodiment, the FUBP1 inhibitor is capable of reducing HBsAg and/or HBeAg in vivo in an HBV infected individual.

FUBP1 Inhibitors for Use in Treatment of HBV

[0136] Without being bound by theory, it is believed that FUBP1 is involved in the stabilization of the cccDNA in the cell nucleus, and by preventing the binding of FUBP1 to DNA, in particular cccDNA, the cccDNA is destabilised and becomes prone to degradation. One embodiment of the invention is therefore a FUBP1 inhibitor which interacts

with the DNA binding domain of FUBP1 protein, and prevents or reduces binding to cccDNA.

[0137] Small molecules inhibiting FUBP1 have been identified in relation to FUBP1's role in cancer, where the small molecule inhibits the DNA binding activity of FUBP1, in particular the binding to the FUSE element on a single stranded DNA. In the present invention FUBP1 inhibitors are envisioned as useful in treating HBV. In particular targeting of such small molecule compounds, e.g. via conjugation or formulation, to the liver may be beneficial in the treatment of HBV.

[0138] Huth et al 2004 J Med. Chem Vol 47 p. 4851-4857 discloses a series of benzoyl anthranilic acid compounds capable of binding to a four tandem K homology (KH) repeat of FUBP1. All the compounds disclosed in Huth et al 2004 are hereby incorporated by reference. In particular the compounds of formula I, II or III shown below were found to be efficient in inhibiting FUBP1 DNA binding activity.

I

II

and
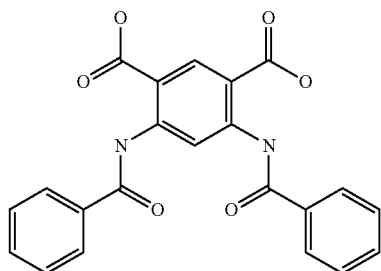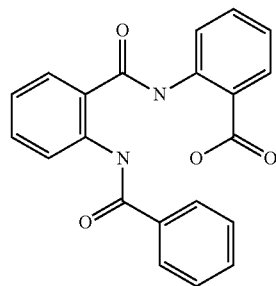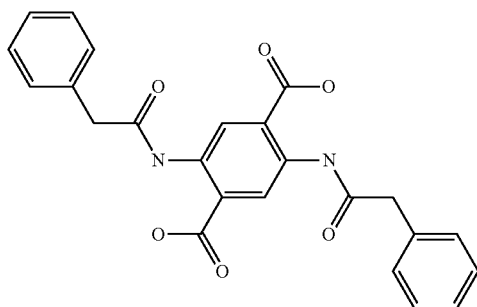
III

[0139] One embodiment of the present invention is a compound of formula I, II or III for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0140] Hauck et al 2016 Bioorganic & Medicinal Chemistry Vol 24 p. 5717-5729 describes an additional series of compounds with high FUBP1 inhibitory potential (see table

2, hereby incorporated by reference). In particular, the following compounds of formula IV were effective in inhibiting FUBP1 activity
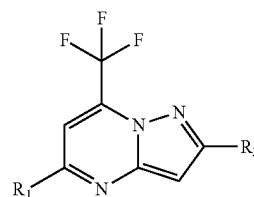
IV

wherein R1 is selected from
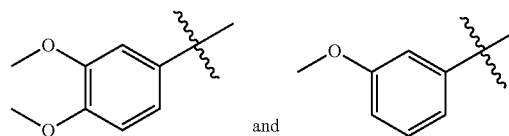
and

and R2 is selected from
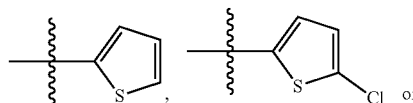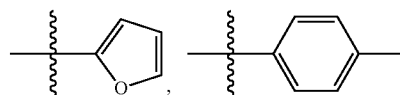
or
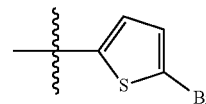
[0141] Specifically the compounds of formula V, VI and VII were shown to have IC50 values below 15 μM.
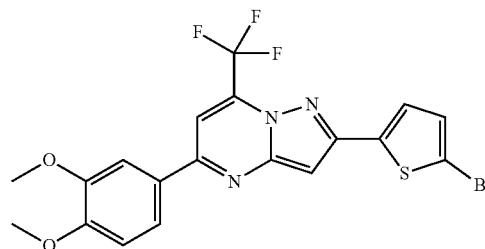
V

2-(5-Bromothiophen-2-yl)-5-(3,4-dimethoxyphenyl)-7-(trifluoromethyl)pyrazolo[1,5-a]pyrimidine

[0142]

VI
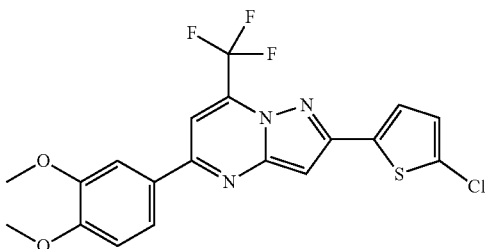
2-(5-Chlorothiophen-2-yl)-5-(3,4-dimethoxyphenyl)-7-(trifluoromethyl)pyrazolo[1,5-a]pyrimidine

[0143]

VII
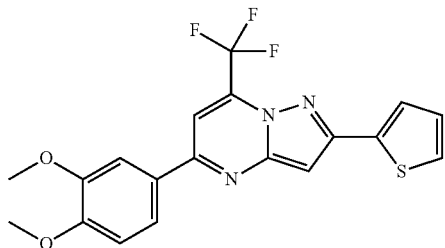
5-(3,4-Dimethoxyphenyl)-2-(thiophen-2-yl)-7-(trifluoromethyl)pyrazolo[1,5-a]pyrimidine

[0144] One embodiment of the present invention is a compound of formula IV for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0145] One embodiment of the present invention is a compound of formula V, VI or VI for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0146] The S-adenosyl-L-methionine (SAM) competitive inhibitor, GSK343 (formula VIII), is currently in preclinical development for osteosarcoma. It has been shown that GSK343 inhibits FUBP1 expression in osteosarcoma cells (Xiong et al 2016 Int J Onc vol 49 p 623).
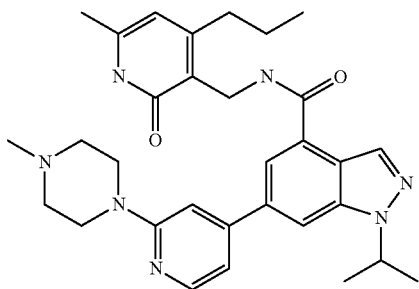
VIII

[0147] One embodiment of the present invention is a compound of formula VII for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0148] The FDA-approved cancer drugs camptothecin (CPT, formula IX) and its derivative SN-38 (7-ethyl-10-hydroxycamptothecin, formula X), which are Topoisomerase I (TOP1) inhibitors, were recently shown to inhibit FUBP1 activity as well by preventing the FUBP1/FUSE interaction (Hosseini et al 2017 Biochemical Pharmacology Vol 146 p. 53-62).
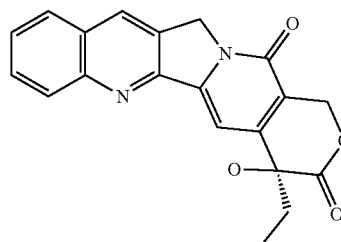
IX

Camptothecin ((+)-4(S)-Ethyl-4-hydroxy-3,4,12,14-tetrahydro-1H-pyrano[3',4':6,7]indolizino[1,2-b]quinoline-3,14-dione)

[0149]
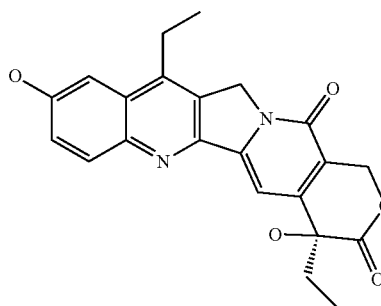
X

SN-38 (4(S),11-Diethyl-4,9-dihydroxy-3,4,12,14-tetrahydro-1H-pyrano[3',4':6,7]indolizino[1,2-b]quinoline-3,14-dione 7-Ethyl-10-hydroxycamptothecin)

[0150] One embodiment of the present invention is a compound of formula IX or X for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0151] Tringali et al 2012 Journal of Pharmacy and Pharmacology Vol 64, p. 360-365 describes the pharmacokinetic profile SN-38 conjugated to hyaluronic acid (HA-SN-38, formula XI) and shows an increased distribution to the liver.

XI
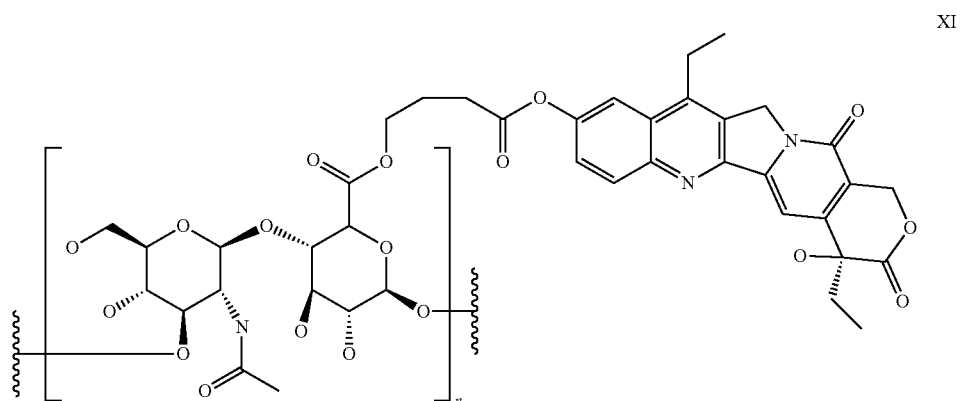
[0152] One embodiment of the present invention is a compound of formula XI for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0153] Various lipid conjugates of SN-38 also exist in the literature. WO2006/082053 for example describes the molecule of formula XII

XII
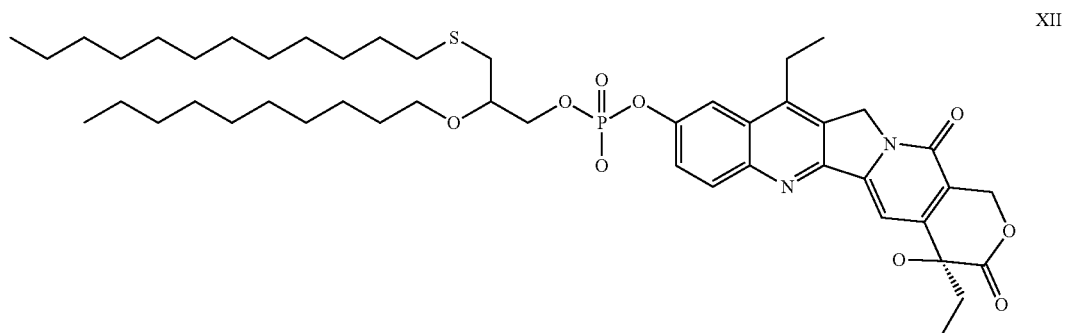
[0154] CN105777770 describes a palmitate conjugated SN-38 shown in formula XIII below.
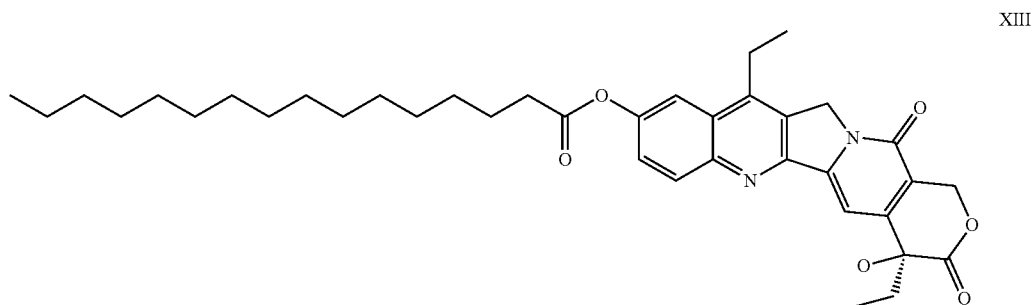
XIII

[0155] One embodiment of the present invention is a compound of formula XII or XIII for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0156] In a further aspect of the invention the FUBP1 inhibitors for use in treatment and/or prevention of Hepatitis B virus (HBV) infection can be targeted directly to the liver by covalently attaching them to a conjugate moiety capable of binding to the asialoglycoprotein receptor (ASGPr), such as divalent or trivalent GalNAc cluster.

Nucleic Acid Molecules of the Invention

[0157] Nucleic acid molecules are potentially excellent FUBP1 inhibitors since they can target the FUBP1 transcript and promote its degradation either via the RNA interference pathway or via RNaseH cleavage. Alternatively, nucleic acid molecules such as aptamers can also act as inhibitors of the DNA binding site of FUBP1 in line with the small molecules described above.

[0158] One aspect of the present invention is a nucleic acid molecule for use in treatment and/or prevention of Hepatitis B virus (HBV) infection. Such nucleic acid molecules can be selected from the group consisting of single stranded antisense, oligonucleotide; siRNA molecule; or shRNA molecule.

[0159] The present section describes novel nucleic acid molecules suitable for use in treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0160] The nucleic acid molecules of the present invention are capable of inhibiting the expression of FUBP1 in vitro and in vivo. The inhibition is achieved by hybridizing an oligonucleotide to a target nucleic acid encoding FUBP1.

[0161] The target nucleic acid may be a mammalian FUBP1 sequence, such as a sequence selected from the group consisting of SEQ ID NO: 1 to 20. It is advantageous if the mammalian FUBP1 sequence is selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7 and 8.

[0162] In some embodiments, the nucleic acid molecule of the invention is capable of modulating the expression of FUBP1 by inhibiting or down-regulating it. Preferably, such modulation produces an inhibition of expression of at least 40% compared to the normal expression level of the target, more preferably at least 50%, 60%, 70%, 80%, 90%, 95% or 98% inhibition compared to the normal expression level of the target. In some embodiments, the nucleic acid molecule of the invention is capable of inhibiting expression levels of FUBP1 mRNA by at least 65% -98%, such as 70% to 95%, in vitro using HepG2-NTCP cells or HBV infected primary human hepatocytes, this range of target reduction is advantageous in terms of selecting nucleic acid molecules with good correlation to the cccDNA reduction. In some embodiments compounds of the invention may be capable of inhibiting expression levels of FUBP1 protein by at least 50% in vitro using HepG2-NTCP cells or HBV infected primary human hepatocytes. The materials and Method section and the Examples herein provide assays which may be used to measure target RNA inhibition in HepG2-NTCP cells or HBV infected primary human hepatocytes as well as cccDNA. The target modulation is triggered by the hybridization between a contiguous nucleotide sequence of the oligonucleotide, such as the guide strand of a siRNA or gapmer region of an antisense oligonucleotide, and the target nucleic acids. In some embodiments, the oligonucleotide of the invention comprises mismatches between the oligonucleotide or the contiguous nucleotide sequence and one or both of the target nucleic acids. Despite mismatches hybridization to the target nucleic acid may still be sufficient to show a desired modulation of FUBP1 expression. Reduced binding affinity resulting from mismatches may advantageously be compensated by increased length of the oligonucleotide and/or an increased number of modified nucleosides capable of increasing the binding affinity to the target within the oligonucleotide sequence. Advantageously, the oligonucleotides of the present invention contain modified nucleosides capable of increasing the binding affinity, such as 2' sugar modified nucleosides, including LNA.

[0163] An aspect of the present invention relates to nucleic acid molecule of 12 to 60 nucleotides in length, which comprises a contiguous nucleotide sequence of 12 to 30 nucleotides in length which is capable of inhibiting the expression of FUBP1.

[0164] In some embodiments, the nucleic acid molecule comprises a contiguous sequence which is at least 90%

complementary, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, or 100% complementary with a region of the target nucleic acid or a target sequence.

[0165] In one embodiment the nucleic acid molecule of the invention, or contiguous nucleotide sequence thereof is fully complementary (100% complementary) to a region of the target nucleic acids, or in some embodiments may comprise one or two mismatches between the oligonucleotide and the target nucleic acids.

[0166] In some embodiments, the nucleic acid molecule comprises a contiguous nucleotide sequence of 12 to 30 nucleotides in length with at least 95% complementary, such as fully (or 100%) complementary, to a target nucleic acid region present in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 and/or SEQ ID NO:4.

[0167] In some embodiments, the nucleic acid molecule or the contiguous nucleotide sequence of the invention is at least 93% complementarity, such as fully (or 100%) complementary, to the target nucleic acid of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO; 7 and /or SEQ ID NO; 8.

[0168] In some embodiments the nucleic acid molecule or the contiguous nucleotide sequence of the invention is at least 95% complementarity, such as fully (or 100%) complementary, to the target nucleic acid of SEQ ID NO: 1 and SEQ ID NO: 5.

[0169] In some embodiments the nucleic acid molecule or the contiguous nucleotide sequence of the invention is at least 95% complementarity, such as fully (or 100%) complementary, to the target nucleic acid of SEQ ID NO: 1, SEQ ID NO: 5 and SEQ ID NO: 9.

[0170] In some embodiments the nucleic acid molecule or the contiguous nucleotide sequence is 100% complementary to position 14200-14218 on SEQ ID NO: 1.

[0171] In some embodiments the nucleic acid molecule or the contiguous nucleotide sequence is 100% complementary to position 14413-14431 on SEQ ID NO: 1.

[0172] In some embodiments the nucleic acid molecule or the contiguous nucleotide sequence is 100% complementary to position 14966-14984 on SEQ ID NO: 1.

[0173] In some embodiments the nucleic acid molecule or the contiguous nucleotide sequence is 100% complementary to position 30344-30362 on SEQ ID NO: 1

[0174] In some embodiments, the nucleic acid molecule of the invention comprises or consists of 12 to 60 nucleotides in length, such as from 13 to 50, such as 14 to 35, such as from 15 to 30 such as from 16 to 22 nucleotides in length.

[0175] In some embodiments, the contiguous nucleotide sequence of the acid molecule which is complementary to the target nucleic acids comprises or consists of 12 to 30, such as from 14 to 25, such as from 16 to 23, such as from 18 to 22, contiguous nucleotides in length.

[0176] In some embodiments, the contiguous nucleotide sequence of the siRNA or shRNA which is complementary to the target nucleic acids comprises or consists of 18 to 28, such as from 19 to 26, such as from 20 to 24, such as from 21 to 23, contiguous nucleotides in length.

[0177] In some embodiments, the contiguous nucleotide sequence of the antisense oligonucleotide which is complementary to the target nucleic acids comprises or consists of 12 to 22, such as from 14 to 20, such as from 16 to 20, such

as from 15 to 18, such as from 16 to 18, such as from 16 to 17 contiguous nucleotides in length.

[0178] It is understood that the contiguous nucleobase sequences (motif sequence) can be modified to for example increase nuclease resistance and/or binding affinity to the target nucleic acid.

[0179] The pattern in which the modified nucleosides (such as high affinity modified nucleosides) are incorporated into the oligonucleotide sequence is generally termed oligonucleotide design. The oligonucleotides of the invention are designed with modified nucleosides and RNA nucleosides (in particular for siRNA and shRNA molecules) or DNA nucleosides (in particular for single stranded antisense oligonucleotides). Advantageously, high affinity modified nucleosides are used.

[0180] In an embodiment, the oligonucleotide comprises at least 1 modified nucleoside, such as at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8 modified nucleosides. In an embodiment the oligonucleotide comprises from 1 to 8 modified nucleosides, such as from 2 to 7 modified nucleosides, such as from 3 to 6 modified nucleosides, such as from 4 to 6 modified nucleosides, such as 4 or 5 modified nucleosides. Suitable modifications are described in the "Definitions" section under "modified nucleoside", "high affinity modified nucleosides", "sugar modifications", "2' sugar modifications" and Locked nucleic acids (LNA)".

[0181] In an embodiment, the oligonucleotide comprises one or more sugar modified nucleosides, such as 2' sugar modified nucleosides. Preferably the oligonucleotide of the invention comprise one or more 2' sugar modified nucleoside independently selected from the group consisting of 2'-O-alkyl-RNA, 2'-O-methyl-RNA, 2'-alkoxy-RNA, 2'-O-methoxyethyl-RNA, 2'-amino-DNA, 2'-fluoro-DNA, arabino nucleic acid (ANA), 2'-fluoro-ANA and LNA nucleosides. It is advantageous if one or more of the modified nucleoside(s) is a locked nucleic acid (LNA). Often used LNA nucleosides are oxy-LNA or cET.

[0182] In a further embodiment the oligonucleotide comprises at least one modified internucleoside linkage. Suitable internucleoside modifications are described in the "Definitions" section under "Modified internucleoside linkage". It is advantageous if at least 2 to 3 internucleoside linkages at the 5' or 3' end of the oligonucleotide are phosphorothioate internucleoside linkages. For single stranded antisense oligonucleotides it is advantageous if at least 75%, such as, such as all, the internucleoside linkages within the contiguous nucleotide sequence are phosphorothioate internucleoside linkages.

[0183] In a further aspect of the invention the nucleic acid molecules, such as the antisense oligonucleotide, siRNA or shRNA, of the invention can be targeted directly to the liver by covalently attaching them to a conjugate moiety capable of binding to the asialoglycoprotein receptor (ASGPr), such as divalent or trivalent GalNAc cluster.

Conjugates

[0184] Since HBV infection primarily affects the hepatocytes in the liver it is advantageous to conjugate the FUBP1 inhibitor to a conjugate moiety that will increase the delivery of the inhibitor to the liver compared to the unconjugated inhibitor. In one embodiment liver targeting moieties are selected from moieties comprising cholesterol or other lipids or conjugate moieties capable of binding to the asialoglycoprotein receptor (ASGPR).

[0185] In some embodiments, the invention provides a conjugate comprising a nucleic acid molecule of the invention covalently attached to a conjugate moiety.

[0186] The asialoglycoprotein receptor (ASGPR) conjugate moiety comprises one or more carbohydrate moieties capable of binding to the asialoglycoprotein receptor (ASPGR targeting moieties) with affinity equal to or greater than that of galactose. The affinities of numerous galactose derivatives for the asialoglycoprotein receptor have been studied (see for example: Jobst, S. T. and Drickamer, K. J B. C. 1996, 271, 6686) or are readily determined using methods typical in the art.

[0187] In one embodiment the conjugate moiety comprises at least one asialoglycoprotein receptor targeting moiety selected from group consisting of galactose, galactosamine, N-formyl-galactosamine, N-acetylgalactosamine, N-propionyl-galactosamine, N-n-butanoyl-galactosamine and N-isobutanoylgalactosamine. Advantageously the asialoglycoprotein receptor targeting moiety is N-acetylgalactosamine (GalNAc).

[0188] To generate the ASGPR conjugate moiety the ASPGR targeting moieties (preferably GalNAc) can be attached to a conjugate scaffold. Generally, the ASPGR targeting moieties can be at the same end of the scaffold. In one embodiment, the conjugate moiety consists of two to four terminal GalNAc moieties linked to a spacer which links each GalNAc moiety to a brancher molecule that can be conjugated to the antisense oligonucleotide.

[0189] In a further embodiment, the conjugate moiety is mono-valent, di-valent, tri-valent or tetra-valent with respect to asialoglycoprotein receptor targeting moieties. Advantageously the asialoglycoprotein receptor targeting moiety comprises N-acetylgalactosamine (GalNAc) moieties.

[0190] GalNAc conjugate moieties can include, for example, those described in WO 2014/179620 and WO 2016/055601 and PCT/EP2017/059080 (hereby incorporated by reference), as well as small peptides with GalNAc moieties attached such as Tyr-Glu-Glu-(aminohexyl Gal-NAc)3 (YEE(ahGalNAc)3; a glycotripeptide that binds to asialoglycoprotein receptor on hepatocytes, see, e.g., Duff, et al., Methods Enzymol, 2000, 313, 297); lysine-based galactose clusters (e.g., L3G4; Biessen, et al., Cardovasc. Med., 1999, 214); and cholane-based galactose clusters (e.g., carbohydrate recognition motif for asialoglycoprotein receptor).

[0191] The ASGPR conjugate moiety, in particular a tri-valent GalNAc conjugate moiety, may be attached to the 3'- or 5'-end of the oligonucleotide using methods known in the art. In one embodiment the ASGPR conjugate moiety is linked to the 5'-end of the oligonucleotide.

[0192] In one embodiment the conjugate moiety is a tri-valent N-acetylgalactosamine (GalNAc), such as those shown in FIG. 1, in particular as shown in FIG. 1D.

Method of Manufacture

[0193] In a further aspect, the invention provides methods for manufacturing the nucleic acid molecule of the invention comprising reacting nucleotide units and thereby forming covalently linked contiguous nucleotide units comprised in the oligonucleotide. Preferably, the method uses phophoramidite chemistry (see for example Caruthers et al, 1987,

Methods in Enzymology vol. 154, pages 287-313). In a further embodiment the method further comprises reacting the oligonucleotide with a conjugating moiety (ligand) to covalently attach the conjugate moiety to the oligonucleotide. In a further aspect a method is provided for manufacturing the composition of the invention, comprising mixing the oligonucleotide or conjugated oligonucleotide of the invention with a pharmaceutically acceptable diluent, solvent, carrier, salt and/or adjuvant. Where the nucleic acid molecule is a siRNA the oligonucleotides are paired and allowed to form double stranded structures.

Pharmaceutical Salt

[0194] The compounds according to the present invention may exist in the form of their pharmaceutically acceptable salts. The term "pharmaceutically acceptable salt" refers to conventional acid-addition salts or base-addition salts that retain the biological effectiveness and properties of the compounds of the present invention and are formed from suitable non-toxic organic or inorganic acids or organic or inorganic bases. Acid-addition salts include for example those derived from inorganic acids such as hydrochloric acid, hydrobromic acid, hydroiodic acid, sulfuric acid, sulfamic acid, phosphoric acid and nitric acid, and those derived from organic acids such as p-toluenesulfonic acid, salicylic acid, methanesulfonic acid, oxalic acid, succinic acid, citric acid, malic acid, lactic acid, fumaric acid, and the like. Base-addition salts include those derived from ammonium, potassium, sodium and, quaternary ammonium hydroxides, such as for example, tetramethyl ammonium hydroxide. The chemical modification of a pharmaceutical compound into a salt is a technique well known to pharmaceutical chemists in order to obtain improved physical and chemical stability, hygroscopicity, flowability and solubility of compounds. It is for example described in Bastin, Organic Process Research & Development 2000, 4, 427-435 or in Ansel, In: Pharmaceutical Dosage Forms and Drug Delivery Systems, 6th ed. (1995), pp. 196 and 1456-1457. For example, the pharmaceutically acceptable salt of the compounds provided herein may be a sodium salt.

[0195] In a further aspect the invention provides a pharmaceutically acceptable salt of the antisense oligonucleotide or a conjugate thereof. In a preferred embodiment, the pharmaceutically acceptable salt is a sodium or a potassium salt.

Pharmaceutical Compositions

[0196] In a further aspect, the invention provides pharmaceutical compositions comprising a nucleic acid molecule and/or conjugate compounds of the invention or salts thereof and a pharmaceutically acceptable diluent, carrier, salt and/or adjuvant. A pharmaceutically acceptable diluent includes phosphate-buffered saline (PBS). In some embodiments the pharmaceutically acceptable diluent is sterile phosphate buffered saline. In some embodiments the oligonucleotide is used in the pharmaceutically acceptable diluent at a concentration of 50 -300 μM solution.

[0197] Suitable formulations for use in the present invention are found in Remington's Pharmaceutical Sciences, Mack Publishing Company, Philadelphia, Pa., 17th ed., 1985. For a brief review of methods for drug delivery, see, e.g., Langer (Science 249:1527-1533, 1990). WO 2007/031091 provides further suitable and preferred examples of pharmaceutically acceptable diluents, carriers and adjuvants (hereby incorporated by reference). Suitable dosages, formulations, administration routes, compositions, dosage forms, combinations with other therapeutic agents, pro-drug formulations are also provided in WO2007/031091.

[0198] Nucleic acid molecules or conjugates of the invention may be mixed with pharmaceutically acceptable active or inert substances for the preparation of pharmaceutical compositions or formulations. Compositions and methods for the formulation of pharmaceutical compositions are dependent upon a number of criteria, including, but not limited to, route of administration, extent of disease, or dose to be administered.

[0199] These compositions may be sterilized by conventional sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile aqueous carrier prior to administration. The pH of the preparations typically will be between 3 and 11, more preferably between 5 and 9 or between 6 and 8, and most preferably between 7 and 8, such as 7 to 7.5. The resulting compositions in solid form may be packaged in multiple single dose units, each containing a fixed amount of the above-mentioned agent or agents, such as in a sealed package of tablets or capsules. The composition in solid form can also be packaged in a container for a flexible quantity, such as in a squeezable tube designed for a topically applicable cream or ointment.

[0200] In some embodiments, the nucleic acid molecule or conjugate of the invention is a prodrug. In particular, with respect to oligonucleotide conjugates the conjugate moiety is cleaved off the oligonucleotide once the prodrug is delivered to the site of action, e.g. the target cell.

Applications

[0201] The nucleic acid molecules of the invention may be utilized as research reagents for, for example, diagnostics, therapeutics and prophylaxis.

[0202] In research, such nucleic acid molecules may be used to specifically modulate the synthesis of FUBP1 protein in cells (e.g. in vitro cell cultures) and experimental animals thereby facilitating functional analysis of the target or an appraisal of its usefulness as a target for therapeutic intervention. Typically, the target modulation is achieved by degrading or inhibiting the mRNA producing the protein, thereby prevent protein formation or by degrading or inhibiting a modulator of the gene or mRNA producing the protein.

[0203] If employing the nucleic acid molecules of the invention in research or diagnostics the target nucleic acid may be a cDNA or a synthetic nucleic acid derived from DNA or RNA.

[0204] Also encompassed by the present invention is an in vivo or in vitro method for modulating FUBP1 expression in a target cell which is expressing FUBP1, said method comprising administering a nucleic acid molecule, conjugate compound or pharmaceutical composition of the invention in an effective amount to said cell.

[0205] In some embodiments, the target cell, is a mammalian cell in particular a human cell. The target cell may be an in vitro cell culture or an in vivo cell forming part of a tissue in a mammal. In preferred embodiments the target cell is present in in the liver. The target cell may be a hepatocyte.

[0206] One aspect of the present invention is related the nucleic acid molecules, conjugate compounds or pharmaceutical compositions of the invention for use as a medicament.

[0207] In an aspect of the invention the nucleic acid molecules, conjugate compound or pharmaceutical composition of the invention is capable of reducing the cccDNA level in the infected cells and therefore inhibiting HBV infection. In particular, the antisense oligonucleotide is capable of affecting one or more of the following parameters i) reducing cccDNA and/or ii) reducing pgRNA and/or iii) reducing HBV DNA and/or iv) reducing HBV viral antigens in an infected cell.

[0208] For example, nucleic acid molecule that inhibits HBV infection may reduce i) the cccDNA levels in an infected cell by at least 40% such as 50%, 60%, 70%, 80%, or 90% reduction compared to controls; or ii) the level of pgRNA by at least 40% such as 50%, 60%, 70%, 80%, or 90% reduction compared to controls. The controls may be untreated cells or animals, or cells or animals treated with an appropriate control.

[0209] Inhibition of HBV infection may be measured in vitro using HBV infected primary human hepatocytes or in vivo for nucleic acid molecules complementary to mouse FUBP1 using HBV minicircle mouse (available at Covance Shanghai, see also Guo et al 2016 Sci Rep 6: 2552 and Yan et al 2017 J Hepatology 66(6):1149-1157) or humanized hepatocytes PXB mouse model (available at PhoenixBio, see also Kakuni et al 2014 Int. J. Mol. Sci. 15:58-74). Inhibition of secretion of HBsAg and/or HBeAg may be measured by ELISA, e.g. by using the CLIA ELISA Kit (Autobio Diagnostic) according to the manufacturers' instructions. Reduction of intracellular cccDNA or HBV mRNA and pgRNA may be measured by qPCR, e.g. as described in the Materials and Methods section. Further methods for evaluating whether a test compound inhibits HBV infection are measuring secretion of HBV DNA by qPCR e.g. as described in WO 2015/173208 or using Northern Blot; in-situ hybridization, or immuno-fluorescence.

[0210] Due to the reduction of FUBP1 levels the nucleic acid molecules, conjugate compounds or pharmaceutical compositions of the present invention can be used to inhibit development of or in the treatment of HBV infection. In particular, the destabilization and reduction of the cccDNA, the nucleic acid molecules, conjugate compounds or pharmaceutical compositions of the present invention more efficiently inhibits development of or treats a chronic HBV infection as compared to a compound that only reduces secretion of HBsAg.

[0211] Accordingly, one aspect of the present invention is related to use of the nucleic acid molecule, conjugate compounds or pharmaceutical compositions of the invention to reduce cccDNA and/or pgRNA in an HBV infected individual.

[0212] A further aspect of the invention relates to the use of the nucleic acid molecules, conjugate compounds or pharmaceutical compositions of the invention to inhibit development of or treat a chronic HBV infection.

[0213] A further aspect of the invention relates to the use of the nucleic acid molecules, conjugate compounds or pharmaceutical compositions of the invention to reduce the infectiousness of a HBV infected person. In a particular aspect of the invention, the nucleic acid molecules, conju-

gate compounds or pharmaceutical compositions of the invention inhibits development of a chronic HBV infection.

[0214] The subject to be treated with the nucleic acid molecules, conjugate compounds or pharmaceutical compositions of the invention (or which prophylactically receives antisense oligonucleotides, conjugate compounds or pharmaceutical compositions of the present invention) is preferably a human, more preferably a human patient who is HBsAg positive and/or HBeAg positive, even more preferably a human patient that is HBsAg positive and HBeAg positive.

[0215] Accordingly, the present invention relates to a method of treating and/or preventing a HBV infection, wherein the method comprises administering an effective amount of the nucleic acid molecules, conjugate compounds or pharmaceutical compositions of the invention.

[0216] The invention also provides for the use of a nucleic acid molecule, a conjugate compound or a pharmaceutical composition of the invention for the manufacture of a medicament, in particular a medicament for use in the treatment or prevention of HBV infection or chronic HBV infection or reduction of the infectiousness of a HBV infected person. In preferred embodiments the medicament is manufactured in a dosage form for subcutaneous administration.

[0217] The invention also provides for the use of a nucleic acid molecule, a conjugate compound, the pharmaceutical composition of the invention for the manufacture of a medicament wherein the medicament is in a dosage form for intravenous administration.

[0218] The nucleic acid molecule, conjugate or the pharmaceutical composition of the invention may be used in a combination therapy. For example, nucleic acid molecule, conjugate or the pharmaceutical composition of the invention may be combined with other anti-HBV agents such as interferon alpha-2b, interferon alpha-2a, and interferon alphacon-1 (pegylated and unpegylated), ribavirin, lamivudine (3TC), entecavir, tenofovir, telbivudine (LdT), adefovir, or other emerging anti-HBV agents such as a HBV RNA replication inhibitor, a HBsAg secretion inhibitor, a HBV capsid inhibitor, an antisense oligomer (e.g. as described in WO2012/145697, WO 2014/179629 and WO2017/216390), a siRNA (e.g. described in WO 2005/014806, WO 2012/024170, WO 2012/2055362, WO 2013/003520, WO 2013/159109, WO 2017/027350 and WO2017/015175), a HBV therapeutic vaccine, a HBV prophylactic vaccine, a HBV antibody therapy (monoclonal or polyclonal), or TLR 2, 3, 7, 8 or 9 agonists for the treatment and/or prophylaxis of HBV.

Administration

[0219] The nucleic acid molecule, conjugate compounds or pharmaceutical composition of the invention is formulated, dosed, and administered in a fashion consistent with good medical practice. Factors for consideration in this context include the particular mammal being treated, the clinical condition of the individual patient, the site of delivery of the agent, the method of administration, the scheduling of administration, the age and sex of the patients and other factors known to medical practitioners. Herein, an "effective amount" (also known as "(therapeutically) effective dose") means the amount of a compound that will elicit the biological or medical response of a subject that is being sought by a medical doctor or other clinician. The "effective

amount" of an nucleic acid molecule, conjugate compound or pharmaceutical composition of the invention, will be governed by such considerations, and is the minimum amount necessary to inhibit HBsAg and/or HBeAg. For example, such amount may be below the amount that is toxic to the cells of the recipient, or to the mammal as a whole.

[0220] In some embodiments, the nucleic acid molecule, conjugate or pharmaceutical composition of the invention is administered at a dose of 0.1-15 mg/kg, such as from 0.2-10 mg/kg, such as from 0.25-5 mg/kg. The administration can be once a week, every 2nd week, every third week or even once a month.

[0221] The nucleic acid molecules, conjugates or pharmaceutical compositions of the present invention may be administered topical (such as, to the skin, inhalation, ophthalmic or otic) or enteral (such as, orally or through the gastrointestinal tract) or parenteral (such as, intravenous, subcutaneous, intra-muscular, intracerebral, intracerebroventricular or intrathecal).

[0222] In a preferred embodiment the nucleic acid molecule, conjugate compounds or pharmaceutical compositions of the present invention are administered by a parenteral route including intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion. In one embodiment the active oligonucleotide or oligonucleotide conjugate is administered intravenously. With GalNAc conjugated compounds it may be advantageous to administer subcutaneously in order to delay saturation of the ASGP receptor.

Combination Therapies

[0223] In some embodiments the oligonucleotide, oligonucleotide conjugate or pharmaceutical composition of the invention is for use in a combination treatment with another therapeutic agent. The therapeutic agent can for example be the standard of care for the diseases or disorders described above.

[0224] By way of example, the oligomer or the oligomer conjugate of the present invention may be used in combination with other actives, such as oligonucleotide-based antivirals—such as sequence specific oligonucleotide-based antivirals—acting either through antisense (including LNA antisense oligonucleotides as described in WO 2011/047312 or WO2015/173208 or MOE antisense oligonucleotides such as GSK3389404 or those described in WO 2012/145697 or WO 2014/79629), siRNAs (such as ARC-520, ARC-521, ARB-1467 or those described in WO2 013/003520, WO2016/077321, US 2017/0016000 or WO 2017/027350), aptamers, morpholinos or any other antiviral, nucleotide sequence-dependent mode of action.

[0225] By way of further example, the oligomer or the oligomer conjugate of the present invention may be used in combination with other actives, such as immune stimulatory antiviral compounds, such as interferon (e.g. pegylated interferon alpha), TLR7 agonists (e.g. GS-9620), TLR8 agonists (e.g. GS-9688) or therapeutic vaccines.

[0226] By way of further example, the oligomer or the oligomer conjugate of the present invention may be used in combination with other actives, such as small molecules, with antiviral activity. These other actives could be, for example, nucleoside/nucleotide inhibitors (eg entecavir or tenofovir disoproxil fumarate), encapsidation inhibitors, entry inhibitors (eg Myrcludex B).

[0227] In certain embodiments, the additional therapeutic agent may be an HBV agent, an Hepatitis C virus (HCV) agent, a chemotherapeutic agent, an antibiotic, an analgesic, a nonsteroidal anti-inflammatory (NSAID) agent, an antifungal agent, an antiparasitic agent, an anti-nausea agent, an anti-diarrheal agent, or an immunosuppressant agent.

[0228] In particular related embodiments, the additional HBV agent may be interferon alpha-2b, interferon alpha-2a, and interferon alphacon-1 (pegylated and unpegylated), ribavirin; an HBV RNA replication inhibitor; a second antisense oligomer; an HBV therapeutic vaccine; an HBV prophylactic vaccine; lamivudine (3TC); entecavir (ETV); tenofovir diisoproxil fumarate (TDF); telbivudine (LdT); adefovir; or an HBV antibody therapy (monoclonal or polyclonal).

[0229] In other particular related embodiments, the additional HCV agent may be interferon alpha-2b, interferon alpha-2a, and interferon alphacon-1 (pegylated and unpegylated); ribavirin; pegasys; an HCV RNA replication inhibitor (e.g., ViroPharma's VP50406 series); an HCV antisense agent; an HCV therapeutic vaccine; an HCV protease inhibitor; an HCV helicase inhibitor; or an HCV monoclonal or polyclonal antibody therapy.

Embodiments of the Invention

[0230] The following embodiments of the present invention may be used in combination with any other embodiments described herein.

[0231] 1. A FUBP1 inhibitor for use in the in the treatment and/or prevention of Hepatitis B virus (HBV) infection.

[0232] 2. The FUBP1 inhibitor for the use of embodiment 1, wherein the FUBP1 inhibitor is administered in an effective amount.

[0233] 3. The FUBP1 inhibitor for the use of embodiment 1 or 2, wherein the HBV infection is a chronic infection.

[0234] 4. The FUBP1 inhibitor for the use of embodiments 1 to 3, wherein the FUBP1 inhibitor is capable of reducing cccDNA and/or pgRNA in an infected cell.

[0235] 5. The FUBP1 inhibitor for the use of any one of embodiments 1 to 4, wherein the FUBP1 inhibitor prevents or reduces the binding of FUBP1 to DNA, such as cccDNA.

[0236] 6. The FUBP1 inhibitor for the use of any one of embodiments 1 to 5, wherein the inhibitor prevents or reduces the FUBP1/FUSE interaction.

[0237] 7. The FUBP1 inhibitor for the use of any one of embodiments 1 to 6, wherein the inhibitor interacts with the DNA binding domain of FUBP1 protein.

[0238] 8. The FUBP1 inhibitor for the use of any one of embodiments 1 to 6, wherein the inhibitor interacts with a FUSE element on cccDNA.

[0239] 9. FUBP1 inhibitor for the use of embodiment 6 to 8 , wherein said inhibitor is a small molecule that specifically prevents or reduces binding of FUBP1 protein to cccDNA.

[0240] 10. The FUBP1 inhibitor for the use of any one of claims 1 to 9, wherein the inhibitor is selected from the compounds of Formula VII, IX or X
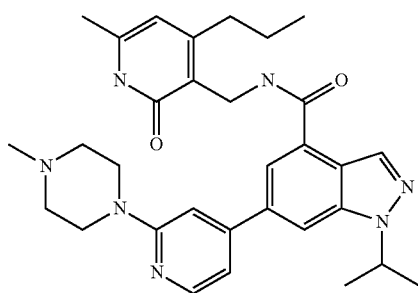
VIII
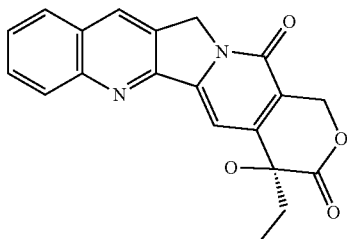
IX
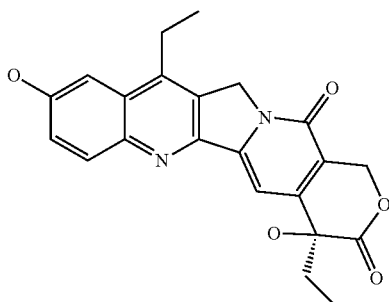
X

[0241] 11. The FUBP1 inhibitor for the use of any one of embodiments 1 to 5, wherein said inhibitor is a genome editing machinery, comprising:

[0242] a. a site-specific DNA nuclease or a poly-nucleotide encoding a site-specific DNA nuclease; and

[0243] b. a guide RNA or a polynucleotide encoding a guide RNA.

[0244] 12. The FUBP1 inhibitor for the use of any one of embodiments 1 to 5, wherein said inhibitor is a nucleic acid molecule comprising or consisting of a contiguous nucleotide sequence of 12 to 30 nucleotides in length which is at least 90% complementary to a mammalian FUBP1 target nucleic acid and capable of reducing the level of the target nucleic acid.

[0245] 13. The FUBP1 inhibitor for the use of embodiment 12, wherein the target nucleic acid is RNA.

[0246] 14. The FUBP1 inhibitor for the use of embodiment 13, wherein the RNA is pre-mRNA.

[0247] 15. The FUBP1 inhibitor for the use of any one of embodiments 11 to14, wherein the nucleic acid molecule is selected from a antisense oligonucleotide, siRNA or shRNA.

[0248] 16. The FUBP1 inhibitor for the use of embodiments 15, wherein the nucleic acid molecule is a single stranded antisense oligonucleotide or a double stranded siRNA.

[0249] 17. The nucleic acid molecule for the use of any one of embodiments 12 to 16, wherein the mammalian FUBP1 target nucleic acid is selected from SEQ ID NO: 1 to 20.

[0250] 18. The nucleic acid molecule for the use of embodiment 12 to 16, wherein the mammalian FUBP1 target nucleic acid is selected from SEQ ID NO: 1 to 8.

[0251] 19. The nucleic acid molecule for the use of embodiment 12 to 16, wherein the mammalian FUBP1 target nucleic acid is selected from SEQ ID NO: 1 to 4.

[0252] 20. The nucleic acid molecule for the use of any one of embodiments 12 to 16, wherein the contiguous nucleotide sequence is at least 98% complementarity to the target nucleic acid of SEQ ID NO: 1 and SEQ ID NO: 5 and SEQ ID NO: 9.

[0253] 21. The nucleic acid molecule for the use of any one of embodiments 12 to 16, wherein the contiguous nucleotide sequence is at least 98% complementarity to the target nucleic acid of SEQ ID NO: 1 and SEQ ID NO: 5.

[0254] 22. The FUBP1 inhibitor for the use of any one of embodiments 1 to 21, wherein the cccDNA in an HBV infected cell is reduced by at least 50%, such as 60%, such as 70%, such as 80%, such 90%, such as 95%, such as 100%, when compared to a control.

[0255] 23. The nucleic acid molecule for the use of any one of embodiments 12 to 22, wherein the FUBP1 mRNA is reduced by at least 50%, such as 60%, such as 70%, such as 80%, such as 90%, such as 95%, such as 100%, when compared to a control.

[0256] 24. A nucleic acid molecule of 12 to 60 nucleotides in length which comprises or consists of a contiguous nucleotide sequence of 12 to 30 nucleotides in length wherein the contiguous nucleotide sequence is at least 90% complementary, such as 95%, such as 98%, such as fully complementarity, to a mammalian FUBP1 target nucleic acid.

[0257] 25. The nucleic acid molecule of embodiment 24, wherein the nucleic acid molecule is chemically produced

[0258] 26. The nucleic acid molecule of embodiment 24 or 25, wherein the mammalian FUBP1 target nucleic acid is selected from SEQ ID NO: 1 to 20.

[0259] 27. The nucleic acid molecule of embodiment 24 or 25, wherein the mammalian FUBP1 target nucleic acid is selected from SEQ ID NO: 1 to 8.

[0260] 28. The nucleic acid molecule of embodiment 24 or 25, wherein the mammalian FUBP1 target nucleic acid is selected from SEQ ID NO: 1 to 4.

[0261] 29. The nucleic acid molecule embodiment 24 or 25, wherein the contiguous nucleotide sequence is at least 98% complementarity to the target nucleic acid of SEQ ID NO: 1 and SEQ ID NO: 5 and SEQ ID NO: 9.

[0262] 30. The nucleic acid molecule of embodiment 24 or 25, wherein the contiguous nucleotide sequence is at least 98% complementarity to the target nucleic acid of SEQ ID NO: 1 and SEQ ID NO: 5.

[0263] 31. The nucleic acid molecule of any one of embodiments 24 to 30, wherein the nucleic acid molecule is 12 to 30 nucleotides in length.

[0264] 32. The nucleic acid molecule of any one of embodiments 24 to 31, wherein the nucleic acid molecule is a double stranded siRNA or a single stranded antisense oligonucleotide.

[0265] 33. The nucleic acid molecule of any one of embodiments 24 to 31, wherein contiguous nucleotide sequence is complementary to a target sequence selected from the group consisting of position 14200-14218, 14413-14431, 14966-14984 and 30344-30362 on SEQ ID NO: 1.

[0266] 34. The nucleic acid molecule of embodiment 24 to 33, which is capable of hybridizing to a target nucleic acid of SEQ ID NO: 1 and SEQ ID NO: 5 with a ΔG° below −15 kcal.

[0267] 35. The nucleic acid molecule of any one of embodiments 24 to 34, wherein the contiguous nucleotide sequence comprises or consists of at least 14 contiguous nucleotides, particularly 15, 16, 17, 18, 19, 20, 21 or 22 contiguous nucleotides.

[0268] 36. The nucleic acid molecule of any one of embodiments 24 to 34, wherein the contiguous nucleotide sequence comprises or consists of from 14 to 22 nucleotides.

[0269] 37. The nucleic acid molecule of embodiment 36, wherein the contiguous nucleotide sequence comprises or consists of from 16 to 18 nucleotides.

[0270] 38. The nucleic acid molecule of embodiment 24 to 37, wherein the nucleic acid molecule comprises or consists of 14 to 25 nucleotides in length.

[0271] 39. The nucleic acid molecule of embodiment 38, wherein the nucleic acid molecule comprises or consists of 16 to 22 nucleotides in length.

[0272] 40. The nucleic acid molecule of any one of embodiments 24 to 39, wherein the contiguous nucleotide sequence has zero to three mismatches compared to the target nucleic acids it is complementary to.

[0273] 41. The nucleic acid molecule of embodiment 40, wherein the contiguous nucleotide sequence has one mismatch compared to the target nucleic acids.

[0274] 42. The nucleic acid molecule of embodiment 40, wherein the contiguous nucleotide sequence has two mismatches compared to the target nucleic acids.

[0275] 43. The nucleic acid molecule of embodiment 40, wherein the contiguous nucleotide sequence is fully complementary to both target nucleic acid sequences.

[0276] 44. The nucleic acid molecule of embodiment 24 to 43, comprising one or more modified nucleosides.

[0277] 45. The nucleic acid molecule of embodiment 44, wherein the one or more modified nucleoside is a high-affinity modified nucleosides.

[0278] 46. The nucleic acid molecule of embodiment 44 or 45, wherein the one or more modified nucleoside is a 2' sugar modified nucleoside.

[0279] 47. The nucleic acid molecule of embodiment 46, wherein the one or more 2' sugar modified nucleoside is independently selected from the group consisting of 2'-O-alkyl-RNA, 2'-O-methyl-RNA, 2'-alkoxy-RNA, 2'-O-methoxyethyl-RNA, 2'-amino-DNA, 2'-fluoro-DNA, 2'-fluoro-ANA and LNA nucleosides.

[0280] 48. The nucleic acid molecule of embodiment 44-47, wherein the one or more modified nucleoside is a LNA nucleoside.

[0281] 49. The nucleic acid molecule of embodiment 48, wherein the modified LNA nucleoside is selected from oxy-LNA, amino-LNA, thio-LNA, cET, and ENA.

[0282] 50. The nucleic acid molecule of embodiment 48 or 49, wherein the modified LNA nucleoside is oxy-LNA with the following 2'-4' bridge —O—CH$_2$—.

[0283] 51. The nucleic acid molecule of embodiment 50, wherein the oxy-LNA is beta-D-oxy-LNA.

[0284] 52. The nucleic acid molecule of embodiment 48 or 49, wherein the modified LNA nucleoside is cET with the following 2'-4' bridge —O—CH(CH$_3$)—.

[0285] 53. The nucleic acid molecule of embodiment 52, wherein the cET is (S)cET, i.e. 6'(S)methyl-beta-D-oxy-LNA.

[0286] 54. The nucleic acid molecule of embodiment 48 or 49, wherein the LNA is ENA, with the following 2'-4' bridge —O—CH$_2$—CH$_2$—.

[0287] 55. The nucleic acid molecule of any one of embodiments 24 to 54, wherein the nucleic acid molecule comprises at least one modified internucleoside linkage.

[0288] 56. The nucleic acid molecule of embodiment 55, wherein the modified internucleoside linkage is nuclease resistant.

[0289] 57. The nucleic acid molecule of embodiment 55 or 56, wherein the modified internucleoside linkages is a phosphorothioate internucleoside linkages.

[0290] 58. The nucleic acid molecule any one of embodiments 24 to 57, wherein the nucleic acid molecule is a double stranded siRNA oligonucleotide capable of interacting with the RISC complex.

[0291] 59. The nucleic acid molecule any one of embodiments 24 to 57, wherein the nucleic acid molecule is an antisense oligonucleotide capable of recruiting RNase H.

[0292] 60. The antisense oligonucleotide of embodiment 59, wherein the antisense oligonucleotide or the contiguous nucleotide sequence is a gapmer.

[0293] 61. The antisense oligonucleotide of embodiment 60, wherein the antisense oligonucleotide or contiguous nucleotide sequence thereof consists of or comprises a gapmer of formula 5'-F-G-F'-3', where region F and F' independently comprise or consist of 1-8 nucleosides, of which 1-4 are 2' sugar modified and defines the 5' and 3' end of the F and F' region, and G is a region between 6 and 16 nucleosides which are capable of recruiting RNaseH.

[0294] 62. The antisense oligonucleotide of embodiment 61, wherein the 2' sugar modified nucleoside independently is selected from the group consisting of 2'-O-alkyl-RNA, 2'-O-methyl-RNA, 2'-alkoxy-RNA, 2'-O-methoxyethyl-RNA, 2'-amino-DNA, 2'-fluoro-DNA, arabino nucleic acid (ANA), 2'-fluoro-ANA and LNA nucleosides.

[0295] 63. The antisense oligonucleotide of embodiment 61 or 62, wherein one or more of the 2' sugar modified nucleosides in region F and F' is a LNA nucleoside

[0296] 64. The antisense oligonucleotide of embodiment 63, wherein all the 2' sugar modified nucleosides in region F and F' are LNA nucleosides.

[0297] 65. The oligonucleotide of embodiment 62 to 64, wherein the LNA nucleoside is selected from beta-D-oxy-LNA, alpha-L-oxy-LNA, beta-D-amino-LNA, alpha-L-amino-LNA, beta-D-thio-LNA, alpha-L-thio-LNA, (S)cET, (R)cET beta-D-ENA and alpha-L-ENA.

[0298] 66. The antisense oligonucleotide of embodiment 62 to 65, wherein region F and F' consist of identical LNA nucleosides.

[0299] 67. The antisense oligonucleotide of embodiment 62 to 66, wherein all the 2' sugar modified nucleosides in region F and F' are oxy-LNA nucleosides.

[0300] 68. The antisense oligonucleotide of any one of embodiments 61 to 67, wherein the nucleosides in region G is DNA and/or alpha-L-LNA nucleosides.

[0301] 69. The antisense oligonucleotide of embodiment 68, wherein region G consists of at least 75% DNA nucleosides.

[0302] 70. The antisense oligonucleotide of embodiment 69, where all the nucleosides in region G are DNA nucleosides.

[0303] 71. A conjugate compound comprising a nucleic acid molecule according to any one of embodiments 24 to 59 or an antisense oligonucleotide according to any one of embodiments 60 to 70, and at least one conjugate moiety covalently attached to said antisense oligonucleotide.

[0304] 72. The conjugate compound of embodiment 71, wherein the nucleic acid molecule is a double stranded siRNA and the conjugate moiety is covalently attached to the sense strand of the siRNA.

[0305] 73. The conjugate compound of embodiment 71 or 72, wherein the conjugate moiety is selected from carbohydrates, cell surface receptor ligands, drug substances, hormones, lipophilic substances, polymers, proteins, peptides, toxins, vitamins, viral proteins or combinations thereof.

[0306] 74. The conjugate compound of any one of embodiments 71 to 73, wherein the conjugate moiety is capable of binding to the asialoglycoprotein receptor.

[0307] 75. The conjugate compound of embodiment 74, wherein the conjugate moiety comprises at least one asialoglycoprotein receptor targeting moiety selected from group consisting of galactose, galactosamine, N-formyl-galactosamine, N-acetylgalactosamine, N-propionyl-galactosamine, N-n-butanoyl-galactosamine and N-isobutanoylgalactosamine.

[0308] 76. The conjugate compound of embodiment 75, wherein the asialoglycoprotein receptor targeting moiety is N-acetylgalactosamine (GalNAc).

[0309] 77. The conjugate compound of embodiment 75 or 76, wherein the conjugate moiety is mono-valent, di-valent, tri-valent or tetra-valent with respect to asialoglycoprotein receptor targeting moieties.

[0310] 78. The conjugate compound of embodiment 77, wherein the conjugate moiety consists of two to four terminal GalNAc moieties and a spacer linking each GalNAc moiety to a brancher molecule that can be conjugated to the antisense compound.

[0311] 79. The conjugate compound of embodiment 78, wherein the spacer is a PEG spacer.

[0312] 80. The conjugate compound of embodiment 74 to 79, wherein the conjugate moiety is a tri-valent N-acetylgalactosamine (GalNAc) moiety.

[0313] 81. The conjugate compound of embodiment 74 to 80, wherein the conjugate moiety is selected from one of the trivalent GalNAc moieties in FIG. 1.

[0314] 82. The conjugate compound of embodiment 81, wherein the conjugate moiety is the trivalent GalNAc moiety in FIG. 1D.

[0315] 83. The conjugate compound of embodiment 71-82, comprising a linker which is positioned between the nucleic acid molecule or the antisense oligonucleotide and the conjugate moiety.

[0316] 84. The conjugate compound of embodiment 83, wherein the linker is a physiologically labile linker.

[0317] 85. The conjugate compound of embodiment 84, wherein the physiologically labile linker is nuclease susceptible linker.

[0318] 86. The oligonucleotide conjugate of embodiment 84 or 85, wherein the physiologically labile linker is composed of 2 to 5 consecutive phosphodiester linkages.

[0319] 87. The conjugate compound of embodiment 74-86, which display improved cellular distribution between liver vs. kidney or improved cellular uptake into the liver of the conjugate compound as compared to an unconjugated nucleic acid molecule or antisense oligonucleotide.

[0320] 88. A pharmaceutical composition comprising a nucleic acid molecule according to any one of embodiments 24 to 59 or an antisense oligonucleotide according to any one of embodiments 60 to 70, a conjugate compound of embodiment 71 to 87 or acceptable salts thereof and a pharmaceutically acceptable diluent, carrier, salt and/or adjuvant.

[0321] 89. A method for identifying a compound that prevents, ameliorates and/or inhibits a hepatitis B virus (HBV) infection, comprising:

[0322] a. contacting a test compound with

[0323] b. FUBP1 polypeptide; or

[0324] c. a cell expressing FUBP1;

[0325] d. measuring the expression and/or activity of FUBP1 in the presence and absence of said test compound; and

[0326] e. identifying a compound that reduces the expression and/or activity FUBP1 and reduces cccDNA.

[0327] 90. An in vivo or in vitro method for modulating FUBP1 expression in a target cell which is expressing FUBP1, said method comprising administering the nucleic acid molecule of any one of embodiments 24 to 59 or an antisense oligonucleotide according to any one of embodiments 60 to 70, a conjugate compound of embodiment 71 to 87 or the pharmaceutical composition of embodiment 88 in an effective amount to said cell.

[0328] 91. The method of embodiments 90, wherein the FUBP1 expression is reduced by at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90%, or at least 95% in the target cell compared to the level without any treatment or treated with a control.

[0329] 92. The method of embodiments 90, wherein the target cell is infected with HBV and the cccDNA in an HBV infected cell is reduced by at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90%, or at least 95% in the HBV infected target cell compared to the level without any treatment or treated with a control.

[0330] 93. A method for treating or preventing a disease comprising administering a therapeutically or prophy-

lactically effective amount of the nucleic acid molecule any one of embodiments 24 to 59 or an antisense oligonucleotide according to any one of embodiments 60 to 70, a conjugate compound of embodiment 71 to 87 or the pharmaceutical composition of embodiment 88 to a subject suffering from or susceptible to the disease.

[0331] 94. The nucleic acid molecule of any one of embodiments 24 to 59 or the antisense oligonucleotide according to any one of embodiments 60 to 70, or the conjugate compound of any one of embodiments 71 to 87 or the pharmaceutical composition of embodiment 88, for use as a medicament for treatment or prevention of a disease in a subject.

[0332] 95. Use of the nucleic acid molecule any one of embodiments 24 to 59 or the antisense oligonucleotide according to any one of embodiments 60 to 70, or the conjugate compound of any one of embodiments 71 to 87 for the preparation of a medicament for treatment or prevention of a disease in a subject.

[0333] 96. The method, the nucleic acid molecule, the antisense oligonucleotide, the conjugate or the use of embodiments 93-95 wherein the subject is a mammal.

[0334] 97. The method, the nucleic acid molecule, the antisense oligonucleotide, the conjugate, or the use of embodiment 96, wherein the mammal is human.

EXAMPLES

[0335] The examples illustrate the invention.

Material and Methods

HepG2-NTCP Cell Line and HBV Infection

[0336] HepG2-NTCP cells (Urban et al. gastroenterology 2014, DOI:10.1053/j.gastro.2013.12.024) were cultured at 37° C. in a humidified atmosphere with 5% CO2 in complete proliferation medium consisting of DMEM+GlutaMAX-I (Gibco #31966-021), 5% HI FCII (Gibco), 1× Pen/Strep (Gibco, #15140), for 2 weeks. For HBV infection HepG2-NTCP cells were trypsinized and resuspended in Infection Medium (complete proliferation medium supplemented with 2.5%

[0337] DMSO) and seeded into a 12 well plate at 400 000 cells/well. After 3 to 4 days, cells were then inoculated with HBV once they reached between 80% to 90% confluency (referred to as day 0) using 500ul Infection Medium plus 4% PEG 8000 and HBV (MOI100) per well. After 16 h the cells were washed 3 times with phosphate-buffered saline and the medium was replaced every 3-4 days with 1 ml Modified Infection Medium (complete proliferation medium supplemented with 3.5% DMSO) per well. HBV genotype D was derived from HepAD38 cell culture supernatant and concentrated using PEG precipitation (Ladner et al 1997 Antimicrobial Agents and Chemotherapy 41(8) 171-1720).

HeLa Cell lines

[0338] HeLa cell line was purchased from European Collection of Authenticated Cell Cultures (ECACC, #93021013) and maintained as recommended by the supplier in a humidified incubator at 37° C. with 5% CO2. For assays, 2,500 cells/well were seeded in a 96 multi well plate in Eagle's Minimum Essential Medium (Sigma, M2279) with 10% fetal bovine serum (FBS), 2 mM Glutamin AQ, 1% NEAA, 25 µg/ml Gentamicin.

Primary Mouse Hepatocytes (PMH)

[0339] Primary mouse hepatocytes were isolated from livers of C57BL/6J mice anesthetized with Pentobarbital after a 2 step perfusion protocol according to the literature (Berry and Friend, 1969, J. Cell Biol; Paterna et al., 1998, Toxicol. Appl. Pharmacol.). The first step was 5 min with HBSS+15 mM HEPES+0.4 mM EGTA followed by 12 min HBSS+20 mM NaHCO3+0.04% BSA (Sigma #A7979)+4 mM CaCL2 (Sigma #21115)+0.2 mg/ml Collagenase Type 2 (Worthington #4176). The Hepatocytes were captured in 5 ml cold Williams medium E (WME) (Sigma #W1878, complemented with 1× Pen/Strep/Glutamine, 10% (v/v) FBS (ATCC #30-2030)) on ice.

[0340] The crude cell suspension was filtered through a 70 µm followed by a 40 µm cell strainer (Falcon #352350 and #352340), filled up to 25 ml with WME and centrifuged at room temperature for 5 min at 50×g to pellet the hepatocytes. The supernatant was removed and the hepatocytes were resuspended in 25 ml WME. After adding 25 ml 90% Percoll solution (Sigma #P4937; pH=8.5-9.5) and centrifugation for 10 min at 25 ° C., 50×g the supernatant and floating cells were removed. To remove the remaining Percoll the pellet was resuspended again in 50 mL WME medium, centrifuged 3 min, 25 ° C. at 50×g and the supernatant discarded. The cell pellet was resuspended in 20 mL WME and cell number and viability determined (Invitrogen, Cellcount) and diluted to 250,000 cells/ml. 25,000 cells/well were seeded on collagen-coated 96-well plates (PD Biocoat Collagen I #356407) and incubated at 37 ° C., 5% CO2. 24 h after seeding the oligonucleotides were added in the desired concentration and the cells were incubated at 37 ° C., 5% CO2 for 72 hours.

HBV Infected PHH Cells

[0341] Fresh primary human hepatocytes (PHH) were provided by PhoenixBio, Higashi-Hiroshima City, Japan (Cat #PXB-cells) in 24-well plate format. Upon arrival the PHH were infected with an MOI of 25 GE using HepG2 2.2.15-derived HBV by incubating the PHH cells with HBV in 4% (v/v) PEG in PHH medium for 16 hours. The cells were then washed thrice with PBS and cultured in a humidified atmosphere with 5% CO2 in fresh PHH medium consisting of DMEM (GIBCO, Cat #21885) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (GIBCO, Cat #10082), 2% (v/v) DMSO, 1% (v/v) Penicillin/Streptomycin (GIBCO, Cat #15140-148), 20 mM HEPES (GIBCO, Cat #15630-080), 44 mM NaHCO3 (Wako, Cat #195-14515), 15 ug/ml L-proline (MP-Biomedicals, Cat #0219472825), 0.25 ug/ml Insulin (Sigma, Cat #I1882), 50 nM Dexamethasone (Sigma, Cat# D8893), 5 ng/ml EGF (Sigma, Cat# E9644), and 0.1 mM L-Ascorbic acid 2-phosphate (Wako, Cat #013-12061).

siRNA Sequences and Compounds

TABLE 4

Human FUBP1 sequences targeted by individual
FUBP1 siRNA molecules

SEQ ID NO:	FUBP1 target sequence	Position on SEQ ID NO: 1	Exon	Individual SIRNA	Dharmacon catalog No.
21	GACAAACCUCUUAGGAUUA	14200-14218	9	FU2	J-011548-06
22	GAGAAGUUCGGAAUGAGUA	14413-14431	10	FU4	J-011548-08
23	GAAAGGAUAGCACAAAUAA	14966-14984	12	FU1	J-011548-05
24	AAUAAGAAGUGGACAAUAC	30344-30362	20	FU3	J-011548-07

[0342] The pool of siRNA (ON-TARGETplus SMART pool siRNA Cat. No. L-011548-00-0005, Dharmacon) contains the four individual siRNA molecules listed in table 4 and are available.

Single Stranded Antisense Oligonucleotides

[0343]

TABLE 4A

List of oligonucleotide motif sequences (indicated by SEQ ID NO),
designs of these, as well as specific oligonucleotide compounds
(indicated by CMP ID NO) designed based on the motif sequence.

SEQ ID NO	Motif sequence	Design	Oligonucleotide compound	CMP ID NO	Start position on SEQ ID NO: 1
34	ataaccatagtcatttga	3-12-3	ATAaccatagtcattTGA	34_1	9138
35	cccataaccatagtcat	3-12-2	CCCataaccatagtcAT	35_1	9142
36	gagccatctacacataaa	2-12-4	GAgccatctacacaTAAA	36_1	10016
37	caagagccatctacacat	3-13-2	CAAgagccatctacacAT	37_1	10019
38	tgtccatttaagaatcca	4-12-2	TGTCcatttaagaatcCA	38_1	10144
39	ttgtgtccatttaagaat	4-12-2	TTGTgtccatttaagaAT	39_1	10147
40	tacctttgctgctgat	3-11-2	TACctttgctgctgAT	40_1	11472
41	gactatacctttgctgc	2-13-2	GActatacctttgctGC	41_1	11476
42	actttgtattcttctgtcat	1-16-3	ActttgtattcttctgtCAT	42_1	12020
43	actttgtattcttctgtc	2-14-2	ACtttgtattcttctgTC	43_1	12022
44	gattcaggtgttccagtta	1-16-2	GattcaggtgttccagtTA	44_1	12393
45	ttcaaacttactggaca	3-11-3	TTCaaacttactggACA	45_1	12412
46	tacatctatcccttcat	2-12-3	TAcatctatcccttCAT	46_1	14452
47	ttacatctatcccttc	2-10-4	TTacatctatccCTTC	47_1	14454
48	cttcctattacaatgccaac	1-17-2	CttcctattacaatgccaAC	48_1	14776
49	atttcttcctattacaatg	2-14-3	ATttcttcctattacaATG	49_1	14781
50	ccatttcttcctattacaa	3-14-2	CCAtttcttcctattacAA	50_1	14783
51	gtttaaaatacattgcc	2-11-4	GTttaaaatacatTGCC	51_1	15537
52	gagtttaaaatacattgcc	2-14-3	GAgtttaaaatacattGCC	52_1	15537

TABLE 4A-continued

List of oligonucleotide motif sequences (indicated by SEQ ID NO),
designs of these, as well as specific oligonucleotide compounds
(indicated by CMP ID NO) designed based on the motif sequence.

SEQ ID NO	Motif sequence	Design	Oligonucleotide compound	CMP ID NO	Start position on SEQ ID NO: 1
53	gctttttatggtttcacc	1-15-2	GcttttatggtttcaCC	53_1	16183
54	atgcttttatggtttcacc	1-17-2	AtgcttttatggtttcaCC	54_1	16183
55	ataatcaacctgtccagct	1-16-2	AtaatcaacctgtccagCT	55_1	23804
56	ataatcaacctgtccagc	1-15-2	AtaatcaacctgtccaGC	56_1	23805
57	ataatcaacctgtccag	1-12-4	AtaatcaacctgtCCAG	57_1	23806
58	tataatcaacctgtccag	2-13-3	TAtaatcaacctgtcCAG	58_1	23806
59	gtataatcaacctgtccag	1-15-3	GtataatcaacctgtcCAG	59_1	23806
60	ataatcaacctgtcca	1-11-4	AtaatcaacctgTCCA	60_1	23807
61	tataatcaacctgtcca	1-12-4	TataatcaacctgTCCA	61_1	23807
62	gtataatcaacctgtcca	1-14-3	GtataatcaacctgtCCA	62_1	23807
63	gtataatcaacctgtcc	3-12-2	GTAtaatcaacctgtCC	63_1	23808
64	tataatcaacctgtcc	2-10-4	TAtaatcaacctGTCC	64_1	23808
65	cccattttcttgtagta	2-13-2	CCcattttcttgtagTA	65_1	23841
66	acccattttcttgtagta	1-15-2	AcccattttcttgtagTA	66_1	23841
67	tacccattttcttgtagta	1-16-2	TacccattttcttgtagTA	67_1	23841
68	atacccattttcttgtagta	1-17-2	AtacccattttcttgtagTA	68_1	23841
69	atacccattttcttgtag	1-14-3	AtacccattttcttgTAG	69_1	23843
70	catacccattttcttgtag	2-15-2	CAtacccattttcttgtAG	70_1	23843
71	catacccattttcttgta	2-14-2	CAtacccattttcttgTA	71_1	23844
72	taaaactaattcaadaat	2-14-2	TGaaactaattcadaaGT	72_1	30264
73	aaatggagctaattcaggag	3-14-3	AAatggagctaattcagGAG	73_1	30265
74	ttgtccacttcttattatt	1-14-4	TtgtccacttcttatTATT	74_1	30341
75	ccccacacaatgaagcaa	2-14-2	CCccacacaatgaagcAA	75_1	30368
76	ttcatcaagtcgtctgcat	1-16-2	TtcatcaagtcgtctgcAT	76_1	30412
77	gatcttcatcaagtcgtc	1-15-2	GatcttcatcaagtcgTC	77_1	30417
78	atattaacctcctatcagt	1-15-3	AtattaacctcctatcAGT	78_1	30511
79	aatattaacctcctatcag	3-13-3	AATattaacctcctatCAG	79_1	30512
80	atttatatcacaaagcatc	2-13-4	ATttatatcacaaagCATC	80_1	30616
81	aagtacatttatatcaca	4-12-2	AAGTacatttatatcaCA	81_1	30623
82	catttattgtaaagcacaaa	2-14-4	CAtttattgtaaagcaCAAA	82_1	30675
83	atcatttattgtaaagca	2-12-4	ATcatttattgtaaAGCA	83_1	30679

Motif sequences represent the contiguous sequence of nucleobases present in the
oligonucleotide.

[0344] Designs refer to the gapmer design, F-G-F', where each number represents the number of consecutive modified nucleosides, e.g. 2' modified nucleosides (first number=5' flank), followed by the number of DNA nucleosides (second number=gap region), followed by the number of modified nucleosides, e.g. 2' modified nucleosides (third number=3' flank), optionally preceded by or followed by further repeated regions of DNA and LNA, which are not necessarily part of the contiguous sequence that is complementary to the target nucleic acid.

[0345] Oligonucleotide compounds represent specific designs of a motif sequence. Capital letters represent beta-D-oxy LNA nucleosides, lowercase letters represent DNA nucleosides, all LNA C are 5-methyl cytosine, and 5-methyl DNA cytosines are presented by "e", all internucleoside linkages are phosphorothioate internucleoside linkages.

siRNA Transfection and Treatment Protocol for HBV Infected Cells

[0346] siRNA treatments were performed in HepG2-NTCP or PHH. Cells were transfected at day 4 and day 6 post infection with 100 ul/well transfection mixture containing 25 nM of either scramble SIRNA (ON-TARGETplus Non-targeting siRNA, D-001810-01, Dharmacon, CO, USA) , FUBP1 pool or single siRNA (SMARTpool: ON-TARGETplus human FUBP1 siRNA, L-011548-00, Dharmacon, CO, USA) or no siRNA (OptiMEM) (0.5 ul/well) diluted in OptiMEM (50 ul/well, Thermo Fisher Scientific Reduced Serum media), and mixed for 5 minutes prior to transfection with Lipofectamine® RNAiMAX Transfection Reagent (3 ul/well) (Thermofisher Scientific catalog No. 13778) diluted in OptiMEM (50 ul/well).

[0347] For PHH southern blot, cells were also treated with an HBV-specific siRNA (HBx siRNA) used as negative control (2 pmol). The HBV-specific siRNA was custom made (GE Life science) (Sense: 5'-GCACUUCGCUUCAC-CUCUG-3' SEQ ID NO: 84) in addition to a second negative control (sicontrol) the Silencer® Negative Control #1 siRNA (Thermo Fisher Scientific). The transfection mixture was added to the cells with 1 ml of DMEM without DMSO and Pen/Strep for 24 h, then medium was replaced with HepG2 NTCP and PHH cell culture medium respectively.

[0348] The cells were harvested at day 8 post infection (after 4 days of siRNA treatment) and stored at –80 ° C. for analysis of FUBP1 mRNA knockdown and the quantification of viral parameters.

Southern Blot Protocol

[0349] For Southern blot, cccDNA purification from HBV infected PHH was conducted using an adapted Hirt extraction protocol previously described in Hirt B 1967 J Mol Biol 26:365-369; and Cai D 2013 Methods Mol Biol. 1030:151-61. Briefly, cells were lysed with HIRT lysis buffer (10 mM Tris pH8, 10 mM EDTA, 0.7% SDS, 2 ul Ambion RNase Mix/ml buffer) to disrupt lipid membranes and viral capsids in order to release all the viral nucleic acids. RNA was removed by RNAse treatment before precipitating down with NaCl (5 M, overnight at 4° C.) the high-molecular-weight cellular chromatin, proteins and protein-associated DNA including the HBV DNA replicative intermediates with viral polymerase covalently attached. After the overnight high-salt protein precipitation, viral DNA (enriched cccDNA) was extracted from the clarified lysates by direct phenol extraction.

[0350] DNA was then quantified and adjusted to load on an agarose gel (1.8% in TAE buffer) 15 µg of each samples in 15 µl per well. DNA ladder (NEB Quick-Load 1 kb), and recombinant HBV were also loaded and the gel was run at 50V for 3 h 30 min.

[0351] For the Southern Blot , the gel was submerged in denaturing buffer (0.5 M NaOH (50 ml), 1.5 M NaCl (150 ml)) and rocked for 30 minutes at room temperature to separate the double-stranded DNA into single DNA strands for later hybridization to the probe. The transfer of DNA from the agarose gel to Hybond-XL membrane (Cat. No: RPN2020S, GE Healthcare) was done using Whatman Nytran SuPerCharge (SPC) TurboBlotter Kit (Cat. No: WHA10416328, Sigma). After transfer, the membrane was cross linked in a UV crosslinker chamber with UV energy dosage at 1800 J. A DIG-labelled RNA probe was generated using HBV plasmid DNA (HBV Genotype D cloned into a pGEM3Z plasmid) and the following primers:

[0352] Forward Primer HBVT7+Fwd 5'-taatacgactcac-tataggggttttttcacctctgcctaatcatc-3' (SEQ ID NO: 89)

[0353] Reverse Primer HBVRev 5'-cctctagagcggccgcaaaaagttgcatggtgctggt-3' (SEQ ID NO: 90) with the DIG RNA Labeling Kit ((SP6/T7) Cat no 11175025910, Roche).

[0354] The membrane was then hybridized with the DIG-labelled probe in DIG Easy Hyb buffer (Cat. No: 11603558001, Roche) at 50° C. overnight while rotating in the hybridization oven. The membrane was then washed and incubated with CDP-Star with NitroBlock minutes before exposing to film. Detection of the signal was done with Fusion Fx (VILBER).

HBV Antigen Measurements

[0355] To evaluate the impact on HBV antigen expression and secretion, supernatants from the cell cultures can be collected. The HBV propagation parameters, HBsAg and HBeAg levels, are measured using CLIA ELISA Kits (Auto-bio Diagnostic #CL0310-2, #CL0312-2), according to the manufacturer's protocol. Briefly, 25 µL of supernatant per well is transferred to the respective antibody coated micro-titer plate and 25 µL of enzyme conjugate reagent is added. The plate is incubated for 60min on a shaker at room temperature before the wells are washed five times with washing buffer using an automatic washer. 25 µL of sub-strate A and B were added to each well. The plates are incubated on a shaker for 10 min at room temperature before luminescence is measured using an Envision luminescence reader (Perkin Elmer).

Real-Time PCR for Intracellular HBV pgRNA and FUBP1 RNA

[0356] FUBP1 RNA and HBV pgRNA were quantified by RTqPCR using SYBR green. Results were normalized over the human Gus B endogenous control. The mRNA expression was analysed using the comparative cycle threshold 2-ΔΔCt method normalized to the reference gene Gus B and to non-treated cells. Primers used for FUBP1 RNA and HBV pgRNA quantification are listed in table 5. RTqPCR conditions were as follow: 48° C., 15 min (RT step); 95° C., 10 min; 40 cycles at 95° C., 15 sec and 60° C., 1 min.

TABLE 5

FUBP1 and HBV pgRNA qPCR primers

Parameter		Primer Sequence	SEQ ID
FUBP1:	Fwd	TGTCCAGTCAGCAAAACGGTT	25
	Rev	GAACTGCATTTCCCGGTCCAT	26
HBV pgRNA	Fwd	GGAGTGTGGATTCGCACTCCT	27
(modified from	Rev	AGATTGAGATCTTCTGCGAC	28
Volz et al.	probe[6FAM]-AGGCAGGTCCCCTA	29
Gastroenterology		GAAGAAGAACTCC-[BHQ1]	
2007)			

Housekeeping gene primers Gus B: Hs99999908_m1 (Thermo Fisher Scientific)

HBV DNA and cccDNA Taqman qPCR

[0357] DNA was extracted from HBV infected cells (HepG2 NTCP or PHH) using the MasterPure™ DNA Purification Kit (Epicentre, Madison, Wisconsin USA) protocol. cccDNA levels were determined after digestion with T5 exonuclease (New England Biolabs, MA, USA) using 10 U of T5 for 500 ng of DNA, 1 hour at 37° C. in 20 ul total volume. After digestion, the samples were diluted to 50 ul of which 4 ul were used for the qPCR reaction. Quantitative real-time polymerase chain reaction measurements were performed on the ViiA7 Real-Time PCR System (Life Technologies). qPCR was performed with the Taqman® Fast Advanced Master Mix (Life Technologies, Cat. No 4444557). qPCR conditions were as follow: 50° C., 2 min; 95° C., 10 min; 40 cycles at 95° C., 15 sec and 60° C., 1 min. cccDNA Primers and probe are shown in table 6.

TABLE 6

cccDNA qPCR primers and probe.

Parameter		Primer Sequence	SEQ ID NO
cccDNA	Fwd	5'-CCGTGTGCACTTCGCTTCA-3'	30
	Rev	5'-GCACAGCTTGGAGGCTTGA-3'	31
	Probe	5'-[6FAM]CATGGAGACCACCGTG AACGCCC[BHQ1]-3'	32

[0358] The cccDNA primer and probes are described in Malmström et al. 2012 PLOS ONE 7(7): e36349.

[0359] b-globin was quantified using the HBB TaqMan® Gene Expression assay (ID Hs00758889_s1, Thermofischer scientific) and total HBV DNA with Taqman primers (Pa03453406_s1, Thermofischer).

Oligonucleotide Synthesis

[0360] Oligonucleotide synthesis is generally known in the art. Below is a protocol which may be applied. The oligonucleotides of the present invention may have been produced by slightly varying methods in terms of apparatus, support and concentrations used.

[0361] Oligonucleotides are synthesized on uridine universal supports using the phosphoramidite approach on an Oligomaker 48 at 1 umol scale. At the end of the synthesis, the oligonucleotides are cleaved from the solid support using aqueous ammonia for 5-16hours at 60° C. The oligonucleotides are purified by reverse phase HPLC (RP-HPLC) or by solid phase extractions and characterized by UPLC, and the molecular mass is further confirmed by ESI-MS.

Elongation of the Oligonucleotide:

[0362] The coupling of β-cyanoethyl-phosphoramidites (DNA-A(Bz), DNA-G(ibu), DNA-C(Bz), DNA-T, LNA-5-methyl-C(Bz), LNA-A(Bz), LNA-G(dmf), or LNA-T) is performed by using a solution of 0.1 M of the 5'-O-DMT-protected amidite in acetonitrile and DCI (4,5-dicyanoimidazole) in acetonitrile (0.25 M) as activator. For the final cycle, a phosphoramidite with desired modifications can be used, e.g. a C6 linker for attaching a conjugate group or a conjugate group as such. Thiolation for introduction of phosphorthioate linkages is carried out by using xanthane hydride (0.01 M in acetonitrile/pyridine 9:1). Phosphordiester linkages can be introduced using 0.02 M iodine in THF/Pyridine/water 7:2:1. The rest of the reagents are the ones typically used for oligonucleotide synthesis.

[0363] For post solid phase synthesis conjugation a commercially available C6 aminolinker phorphoramidite can be used in the last cycle of the solid phase synthesis and after deprotection and cleavage from the solid support the aminolinked deprotected oligonucleotide is isolated. The conjugates are introduced via activation of the functional group using standard synthesis methods.

Purification by RP-HPLC:

[0364] The crude compounds are purified by preparative RP-HPLC on a Phenomenex Jupiter C18 10μ 150×10 mm column. 0.1 M ammonium acetate pH 8 and acetonitrile is used as buffers at a flow rate of 5 mL/min. The collected fractions are lyophilized to give the purified compound typically as a white solid.

Abbreviations:

[0365] DCI: 4,5-Dicyanoimidazole

[0366] DCM: Dichloromethane

[0367] DMF: Dimethylformamide

[0368] DMT: 4,4'-Dimethoxytrityl

[0369] THF: Tetrahydrofurane

[0370] Bz: Benzoyl

[0371] Ibu: Isobutyryl

[0372] RP-HPLC: Reverse phase high performance liquid chromatography

T_m Assay:

[0373] Oligonucleotide and RNA target (phosphate linked, PO) duplexes are diluted to 3 mM in 500 ml RNase-free water and mixed with 500 ml 2× T_m-buffer (200 mM NaCl, 0. 2mM EDTA, 20 mM Naphosphate, pH 7.0). The solution is heated to 95° C. for 3 min and then allowed to anneal in room temperature for 30 min. The duplex melting temperatures (T_m) is measured on a Lambda 40 UV/VIS Spectrophotometer equipped with a Peltier temperature programmer PTP6 using PE Templab software (Perkin Elmer). The temperature is ramped up from 20° C. to 95° C. and then down to 25° C., recording absorption at 260 nm. First derivative and the local maximums of both the melting and annealing are used to assess the duplex T_m.

Example 1: Chromatin Immunoprecipitation (ChiP) of FUBP1 in HBV Infected HepG2-NTCP

[0374] In this experiment, it was investigated whether FUBP1 binds to cccDNA. In brief, chromatin fragments from HBV infected cells were immunoprecipitated with a

FUBP1-specific antibody and it was analysed whether cccDNA was pulled down together with the FUBP1 Ab.

[0375] HBV infected HepG2-NTCP cells (see materials and method section) were crossed link with 1% formaldehyde in Williams medium without serum or antibiotics for 10 minutes and the reaction was stopped by adding 0.125 M glycine in PBS 1×, 5 min at room temperature. Cells were washed 2 times with cold PBS and harvested with a cell scraper. The pelleted cells were collected after centrifugation at 4500 rpm for 10 min at 4° C. in the presence of 1 ml of PBS 1× with PIC 1× and PMSF 1 mM and stored at −80° C. before proceeding to the chromatin immunoprecipitation. Nuclear extraction was performed using a dounce (tight pestle) after 30 minutes of incubation on ice in Nuclei Lysis buffer (PIPES 5 mM, KCl 85 mM, NP-40 0.5%, PMSF 1 mM and PIC 1×)

[0376] Isolated cross-linked nuclei were sheared by sonication in a 1% SDS sonication buffer to generate cellular chromatin fragments of 200-500 bp. Chromatin amounts from $2×10^6$ cells was separated, 1/10 of this was saved to serve as reference input sample and the reminder was subjected to immunoprecipitation with antibody for 14-16 hours at 4° C. (ChIP is further described in Pollicino et al 2006 Gastroenterology130:823-837). The antibody used for the immunoprecipitation was 5 μg FUBP1-specific Ab (C15410233 Diagenode). After the reverse cross-linking step, the precipitated chromatin sample and the input sample were subjected to Phenol: Chlorophorm: Isoamyl Alcohol (25:24:1) DNA extraction and cccDNA was then quantified by qPCR (see materials and method section on HBV cccDNA quantification starting with T5 exonuclease treatment and with the change that the cccDNA is not normalized to b-globulin (not precipitated by the antibody), but is expressed as the percentage of cccDNA in the input sample).

[0377] cccDNA amount in the samples precipitated with and without FUBP Ab is expressed as % of cccDNA in the input sample (non-ChIP treated sample). The results are shown in table 7. The values represent the mean of three independent experiments.

TABLE 7

cccDNA immunoprecipitated with a FUBP1-specific antibody or control.		
	% cccDNA of input sample	
	Mean	SD
NoAb	0.0043	0.0025
FUBP1	0.1528	0.0364

[0378] The ChIP experiment showed significant enrichment of FUBP1 protein bound to cccDNA compared to negative control (NoAb).

Example 2: Effect of siRNAs Targeting FUBP1 on HBV Parameters in HBV Infected HepG2-NTCP Cells

[0379] In the following experiment, the effect of FUBP1 knock-down on the HBV parameters, HBV pgRNA and cccDNA, was tested.

[0380] HBV infected HepG2-NTCP cells were treated with the pool of siRNAs from Dharmacon (L-011548-00) as described in the Materials and Method section "siRNA transfection and treatment protocol for HBV infected cells".

[0381] Following the 4 days-treatment, FUBP1 mRNA, cccDNA and pgRNA were measured by qPCR. FUBP1 mRNA and pgRNA RT-qPCR was described in the materials and method section "Real-time PCR for intracellular HBV pgRNA and FUBP1 RNA". For cccDNA absolute quantification in HepG2 NTCP cells, b-globin was measured using 500ng non-digested DNA in a total volume of 50 ul of which 4 ul were used in the qPCR and used for normalization. Commercial genomic DNA was used for the standard curve for b-globin and an HBV plasmid DNA (HBV Genotype D cloned into a pGEM3Z plasmid) was used for cccDNA standard curve. The cccDNA primers used were those described in the Material and Method section "HBV DNA and cccDNA taqman qPCR". The results are shown in table 8.

TABLE 8

effect on HBV parameters following knockdown of FUBP1 with pool of siRNA					
	FUBP1 mRNA* (n = 3)		HBV pgRNA* (n = 4)		cccDNA/ b-globin (n = 4)^
	Mean	SD	Mean	SD	Mean	SD
No treatment	1.22	0.80	2.85	1.46	0.08	0.07
siRNA control	1.13	0.11	2.82	0.55	0.10	0.08
FUBP1 siRNA	0.25	0.27	1.12	0.55	0.03	0.04

*FUBP1 and HBV pgRNA were normalized to housekeeping gene Gus B.
^siRNA control was n = 3

[0382] From this it can be seen that the FUBP1 siRNA pool is capable of reducing FUBP1 mRNA as well as HBV pgRNA quite efficiently. A trend showing reduction of cccDNA is also observed, despite the short treatment time.

Example 3: Effect of siRNAs Targeting FUBP1 on HBV Parameters in HBV Infected PHH Cells

[0383] In the following experiment, the effect of FUBP1 knock-down on the HBV parameters, HBV pgRNA, cccDNA and HBV DNA, was tested in HBV infected primary human hepatocytes (PHH),

[0384] HBV infected PHH cells were treated with the pool of siRNAs used in Example 2 (Dharmacon L-011548-00) as well as four individual siRNA compounds FU1, FU2, FU3 and FU4 listed in the siRNA sequence and compound section of the materials and methods. The treatment was conducted as described in the materials and method section "siRNA transfection and treatment protocol for HBV infected cells".

[0385] Following the 4 days-treatment, FUBP1 mRNA, cccDNA and pgRNA were measured by qPCR in technical duplicate. FUBP1 mRNA and pgRNA RT-qPCR was described in the Materials and Method section "Real-time PCR for intracellular HBV pgRNA and FUBP1 RNA". HBV DNA was measured as described in the Materials and Methods section "HBV DNA and cccDNA taqman quantification qPCR" The HBV DNA was normalized to the housekeeping gene b-globin using the indicated Taqman primers. For cccDNA quantification, SYBR green primers for cccDNA and mitochondrial DNA (MitoDNA, house-

keeping control) were performed using Fast SYBR™ green master mix (Thermo Fischer) and the following primers:

```
cccDNA:
Fwd
                          (SEQ ID NO: 85)
5'-CGTCTGTGCCTTCTCATCTGC-3';

Rev
                          (SEQ ID NO: 86)
5'-GCACAGCTTGGAGGCTTGAA-3'

MitoDNA:
Fwd
                          ((SEQ ID NO: 87)
5'-CCGTCTGAACTATCCTGCCC-3';

Rev
                          (SEQ ID NO: 88)
5'-GCCGTAGTCGGTGTACTCGT-3'
```

[0386] qPCR conditions: 95° C. 5 min, then 45 cycles of 95° C.—1 sec, 60° C.—35 sec

[0387] The results are shown in table 9 as the relative expression to non-treated cells using the comparative cycle threshold 2-ΔΔCt method (i.e. the lower the number the larger the target reduction).

TABLE 9

effect on HBV parameters following
knockdown of FUBP1 with siRNA targeting FUBP1

	FUBP1 mRNA[#]		HBV pgRNA[#]		cccDNA*		HBV DNA**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Non-treated	1.030	0.105	1.00	0.07	1.00	0.04	1.00	0.07
FUBP1 siRNA pool	0.66	0.04	0.61	0.04	0.64	0.06	0.70	0.04
FU1	0.68	0.02	0.53	0.10	0.62	0.04	0.51	0.04
FU2	0.72	0.02	0.43	0.06	0.61	0.11	0.96	0.03
FU3	0.60	0.01	0.45	0.02	0.70	0.03	0.94	0.03
FU4	0.71	0.02	0.38	0.01	0.78	0.09	0.61	0.03

*cccDNA was normalized to mitochondrial DNA reference gene
**HBV DNA was normalized to b-globin housekeeping gene
[#]FUBP1 mRNA and HBV pgRNA was normalized to Gus B housekeeping gene.

[0388] From this it can be seen that the FUBP1 siRNA pool as well as the individual siRNA compounds are capable of reducing FUBP1 mRNA and cccDNA by 30 to 40%. The HBV pgRNA is reduced even further. HBV DNA reduction is observed for the pool of siRNA's and the individual siRNA's FU1 and FU4. The lack of reduction observed with individual siRNA's FU2 and FU3 may be due to the short treatment times since HBV DNA reduction generally require longer treatment times.

[0389] The ability of the FUBP1 siRNA pool to reduce cccDNA was confirmed by Southern blot and compared with an HBV targeting siRNA (HBx) which is known not to reduce cccDNA.

[0390] HBV infected PHH cells were treated as described in Materials and Method section "siRNA transfection and treatment protocol for HBV infected cells" and the Southern blot was performed as described in the Materials and Methods section "Southern Blot protocol".

[0391] The Southern blot is shown in FIG. 3 and confirms the qPCR results in this example, since it is clearly seen that the Pool of FUBP1 targeting siRNA's are capable of reducing the cccDNA band whereas the HBV targeting siRNA

does (as expected) not result in any reduction of cccDNA when compared to the control.

Example 4: The Effect of FUBP1 Small Molecule Inhibitors on HBV Parameters in HBV Infected PHH Cells

[0392] Several small molecules have been suggested to affect FUBP1 binding to the FUSE element on DNA. In the following we tested the effect of SN-38 (formula X herein) and GSK343 (formula VIII herein) for their ability to affect HBV parameters, HBV pgRNA, cccDNA and HBV DNA as well as FUBP1 mRNA levels.

[0393] The treatments were performed in HBV infected PHH cells (See Materials and Methods) with small molecules SN-38 (Toronto research chemical, TRC-S589960) and GSK343 (MedChemExpress, MCE-HY-13500) used at 1 μM. The treatment was performed twice weekly starting day 7 and up to day 25 post infection (i.e. treatment was performed on day 7, 11, 14, 18, 21 and 25 post infection). The cells were harvested at day 12 and day 25 post infection (5 days and 18 days of treatment). Following the 5 days or 18 days treatment, FUBP1 mRNA, cccDNA, pgRNA and HBV DNA were measured by qPCR in technical duplicate. FUBP1 mRNA and pgRNA RT-qPCR was measured as described in the materials and method section "Real-time PCR for intracellular HBV pgRNA and FUBP1 RNA". HBV DNA and cccDNA was normalized to b-globin housekeeping gene as described in the Material and Methods section "HBV DNA and cccDNA taqman qPCR".

[0394] The results are shown in table 10A and 10B as the relative expression to non-treated cells (i.e. the lower the number the larger the target reduction). The toxicity of the compounds was also tested using a standard CCK8 toxicity assay, and no cell toxicity was observed for the two compounds (data not shown).

TABLE 10A

Effect on HBV parameters after 5 days treatment with FUBP1
small molecule inhibitors in HBV infected PPH.

	FUBP1 mRNA*		HBV pgRNA[#]		cccDNA*		HBV DNA*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Non-treated	1.00	0.09	1.00	0.05	1.01	0.17	1.01	0.06
SN-38 1 uM	1.04	0.01	0.54	0.03	0.60	0.10	0.42	0.00
GSK343 1 uM	0.90	0.01	0.99	0.02	1.18	0.22	1.05	0.07

TABLE 10B

Effect on HBV parameters after 18 days treatment with FUBP1
small molecule inhibitors in HBV infected PPH

	FUBP1 mRNA[#]		HBV pgRNA[#]		cccDNA*		HBV DNA*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Non-treated	1.00	0.06	1.00	0.00	1.23	0.95	1.04	0.32
SN-38 1 uM	0.96	0.06	0.89	0.15	0.84	0.02	0.51	0.01
GSK343 1 uM	1.27	0.17	1.34	0.09	0.34	0.16	1.29	0.34

*cccDNA HBV DNA was normalized to b-globin housekeeping gene
[#]FUBP1 mRNA and HBVpgRNA were normalized to Gus B housekeeping gene

[0395] From these data it can be seen that the FUBP1 small molecule inhibitors do not affect FUBP1 mRNA,

which is expected since they act as antagonists of the FUBP1 protein. SN-38 has a significant effect on all three HBV parameters at day 12 at day 25 the effects on pgRNA and cccDNA seem to be rebounding. GSK343 does not have a marked effect on pgRNA and HBV DNA, the reduction of cccDNA is however more than 65% at day 25, indicating that this compound requires longer exposure to be effective, but then it is very effective on cccDNA. It remains to be shown if an even longer exposure would also lead to effects on pgRNA and HBVDNA.

Example 5: The Ability of Single Stranded LNA Antisense Oligonucleotides to Reduce FUBP1 In Vitro in Hela Cells

[0396] The antisense oligonucleotides listed in table 4a were tested for their ability to reduce FUBP1 in Hela cells.

[0397] Hela cells were cultured as described in the Materials and Method section. The cells were incubated for 24 hours before addition of oligonucleotides dissolved in PBS. Final concentration of oligonucleotides was 5 and 25 μM, the final culture volume was 100 μl/well. The cells were harvested 3 days after addition of oligonucleotide compounds. RNA was extracted using RNeasy 96 extraction kit (Qiagen,) and followed by one-step RT-QPCR (Quanta Bioscience, qScript XLT 1-Step RT-qPCR ToughMix) using TaqMan assays for the target genes (FUBP1: Hs.PT.58. 26883775 FAM (IDT)) and a house keeping gene (GUSB 4326320E (Thermo Fisher)) according to the manufacturer's protocols.

[0398] The relative FUBP1 mRNA expression levels are shown in table 11 as % of average control samples (PBS-treated cells) i.e. the lower the value the larger the inhibition.

TABLE 11

in vitro efficacy of anti-FUBP1 compounds (single experiment with duplex QPCR). FUBP1 mRNA levels are normalized to GUSB in Hela cells and shown as % of control (PBS treated cells).

| CMP ID NO | % FUBP1 mRNA of control | | | | |
| | 5 μM | | 25 μM | | |
	Avg	sd	Avg	sd	Compound (CMP)
34_1	65.5	1.4	35.2	2.0	ATAaccatagtcattTGA
35_1	58.5	0.1	27.4	0.7	CCCataaccatagtcAT
36_1	81.6	0.1	57.5	4.9	GAgccatctacacaTAAA
37_1	85.7	2.0	64.6	0.9	CAAgagccatctacacAT
38_1	74.8	4.1	49.9	0.3	TGTCcatttaagaatcCA
39_1	76.2	0.9	61.2	1.4	TTGTgtccatttaagaAT
40_1	78.1	2.7	64.2	3.1	TACctttgctgctgAT
41_1	88.2	3.9	66.0	0.8	GActatacctttgctGC
42_1	82.7	2.9	63.2	0.5	Actttgtattcttctgt CAT
43_1	76.3	1.4	56.3	0.8	ACtttgtattcttctgTC
44_1	89.5	3.9	69.9	4.9	Gattcaggtgttccagt TA

TABLE 11-continued

in vitro efficacy of anti-FUBP1 compounds (single experiment with duplex QPCR). FUBP1 mRNA levels are normalized to GUSB in Hela cells and shown as % of control (PBS treated cells).

| CMP ID NO | % FUBP1 mRNA of control | | | | |
| | 5 μM | | 25 μM | | |
	Avg	sd	Avg	sd	Compound (CMP)
45_1	87.3	4.5	66.3	5.4	TTCaaacttactggACA
46_1	81.1	3.0	60.3	0.1	TAcatctatcccttCAT
47_1	83.8	0.5	52.7	1.7	TTacatctatccCTTC
48_1	74.9	12.7	64.1	2.7	Cttcctattacaatgcca AC
49_1	81.3	3.3	68.7	2.2	ATttcttcctattacaATG
50_1	60.9	9.3	43.2	5.7	CCAtttcttcctattacAA
51_1	69.6	3.1	50.5	2.4	GTttaaaatacatTGCC
52_1	68.5	0.9	53.3	1.0	GAgtttaaaatacattGCC
53_1	69.6	1.6	46.8	1.2	Gcttttttatggtttca CC
54_1	77.5	1.6	65.9	3.2	Atgcttttttatggtttca CC
55_1	86.8	3.3	77.8	2.7	Ataatcaacctgtccag CT
56_1	85.0	6.1	82.6	0.0	Ataatcaacctgtcca GC
57_1	84.2	0.9	71.8	1.0	Ataatcaacctgt CCAG
58_1	82.1	1.3	73.3	1.2	TAtaatcaacctgtcCAG
59_1	81.4	0.6	71.2	0.1	Gtataatcaacctgtc CAG
60_1	86.2	1.6	64.8	2.1	Ataatcaacctgt TCCA
61_1	86.0	1.3	66.4	1.5	Tataatcaacctg TCCA
62_1	79.3	2.5	59.5	0.6	Gtataatcaacctgt CCA
64_1	88.9	7.2	56.6	1.5	GTAtaatcaacctgt CC
63_1	68.8	9.1	33.7	1.0	TAtaatcaacct GTCC
65_1	63.0	1.2	37.7	1.8	CCcattttcttgtag TA
66_1	87.4	1.4	67.9	1.9	Acccattttcttgtag TA
67_1	88.8	3.7	81.2	1.1	Tacccattttcttgtag TA
68_1	94.7	3.1	84.9	0.3	Atacccattttcttgtag TA
69_1	87.1	1.8	64.6	0.1	Atacccattttcttgt TAG
70_1	83.6	6.8	81.3	0.7	CAtacccattttcttgt AG
71_1	90.2	4.9	85.6	2.4	CAtacccattttcttg TA
72_1	76.8	1.9	62.8	1.5	TGgagctaattcagga GT
73_1	82.6	1.1	66.9	0.5	AAtggagctaattcag GAG
74_1	73.3	0.8	67.5	2.5	Ttgtccacttcttat TATT

TABLE 11-continued

in vitro efficacy of anti-FUBP1 compounds
(single experiment with duplex QPCR). FUBP1
mRNA levels are normalized to GUSB in Hela
cells and shown as % of control (PBS
treated cells).

CMP ID NO	% FUBP1 mRNA of control				
	5 µM		25 µM		
	Avg	sd	Avg	sd	Compound (CMP)
75_1	75.9	1.0	47.2	0.9	CCccacacaatgaagcAA
76_1	75.2	0.1	53.0	3.4	TtcatcaagtcgtctgcAT
77_1	55.2	3.9	35.2	2.2	GatcttcatcaagtcgTC
78_1	83.5	3.6	69.2	0.5	AtattaacctcctatcAGT
79_1	66.5	0.4	45.7	0.2	AATattaacctcctatCAG
80_1	63.0	2.7	30.2	3.0	ATttatatcacaaagCATC
81_1	61.2	1.3	36.9	2.5	AAGTacatttatatcaCA
82_1	84.6	0.9	68.5	0.5	CAtttattgtaaagcaCAAA
83_1	61.7	0.5	40.4	1.8	ATcatttattgtaaAGCA

[0399] From these data is can be seen that 80% of the single stranded antisense oligonucleotides tested are capable of reducing FUBP1 by at least 30% at the 25 µM dose.

Example 6: The ability of Single Stranded LNA Antisense Oligonucleotides to Reduce FUBP1 In Vitro in PMH Cells

[0400] The oligonucleotides screened in Example 5 also target mouse FUBP1. To verify that the reduction of FUBP1 observed in the human Hela cells also translates to hepatocytes, which is the cell type to be targeted in the treatment of HBV, the same library was tested in primary mouse hepatocytes (PMH).

[0401] The screening in PMH cells was conducted as described in the "Materials and Methods" section under "Primary mouse Hepatocytes" using 5 M and 25 µM oligonucleotide, the final culture volume was 100 µl/well.

[0402] RNA isolation and qPCR to measure FUBP1 mRNA expression levels were conducted as described in example 5 using the following mouse specific primers instead Fubp1: Mm.PT.58.7603777 FAM-MGB (IDT), and a house keeping gene (GusB Mm_01197698_m1 VIC-MGB (IDT) (Thermo Fisher).

[0403] The relative FUBP1 mRNA expression levels are shown in table 12 as % of average control samples (PBS-treated cells) i.e. the lower the value the larger the inhibition.

TABLE 12

in vitro efficacy of anti-FUBP1 compounds
(single experiment with duplex QPCR). FUBP1
mRNA levels are normalized to GUSB in Hela
cells and shown as % of control (PBS
treated cells).

CMP ID NO	% FUBP1 mRNA of control				
	5 µM		25 µM		
	Avg	sd	Avg	sd	Compound (CMP)
34_1	46.1	2.5	25.2	0.5	ATAaccatagtcattTGA
35_1	48.0	0.1	29.0	1.1	CCCataaccatagtcAT
36_1	32.7	0.7	52.9	1.6	GAgccatctacacaTAAA
37_1	50.0	1.4	48.2	0.4	CAAgagccatctacacAT
38_1	48.4	1.6	35.4	0.7	TGTCcatttaagaatcCA
39_1	41.6	0.7	41.5	1.5	TTGTgtccatttaagaAT
40_1	59.2	1.1	52.8	0.5	TACctttgctgctgAT
41_1	88.5	1.1	59.6	0.9	GActataccttgctGC
42_1	75.8	2.0	66.9	1.9	Actttgtattcttctgt CAT
43_1	71.0	2.1	59.4	0.5	ACtttgtattcttctgTC
44_1	71.4	1.4	56.7	1.2	Gattcaggtgttccagt TA
45_1	60.5	1.7	41.1	1.0	TTCaaacttactggACA
46_1	57.5	0.3	41.5	1.2	TAcatctatcccttCAT
47_1	44.5	0.5	38.3	0.5	TTacatctatccCTTC
48_1	58.7	0.1	45.2	0.3	Cttcctattacaatgccа AC
49_1	65.0	3.6	48.2	1.4	ATttcttcctattacaATG
50_1	11.7	3.2	9.0	2.6	CCAtttcttcctattacAA
51_1	64.5	0.8	48.1	0.9	GTttaaaatacatTGCC
52_1	68.1	1.2	51.9	1.5	GAgtttaaaatacattGCC
53_1	46.1	1.3	33.2	0.5	Gctttttatggtttca CC
54_1	62.3	1.6	46.9	2.3	Atgctttttatggtttca CC
55_1	77.8	0.1	66.3	1.6	Ataatcaacctgtccag CT
56_1	78.0	2.9	60.4	1.7	Ataatcaacctgtcca GC
57_1	51.1	1.7	29.2	0.8	AtaatcaacctgtCCAG
58_1	53.7	2.6	37.7	0.4	TAtaatcaacctgtcCAG
59_1	62.8	0.1	37.0	2.1	Gtataatcaacctgtc CAG
60_1	51.6	0.3	40.3	0.8	AtaatcaacctgTCCA
61_1	55.3	2.0	38.1	0.8	Tataatcaacctg TCCA
62_1	38.5	0.9	26.4	0.1	Gtataatcaacctgt CCA
64_1	58.2	2.4	39.7	0.8	GTAtaatcaacctgtCC
63_1	24.7	1.0	12.6	0.4	TAtaatcaacctGTCC
65_1	29.8	0.8	19.1	0.3	CCcattttcttgtagTA

TABLE 12-continued

in vitro efficacy of anti-FUBP1 compounds
(single experiment with duplex QPCR). FUBP1
mRNA levels are normalized to GUSB in Hela
cells and shown as % of control (PBS
treated cells).

CMP ID NO	% FUBP1 mRNA of control				
	5 µM		25 µM		
	Avg	sd	Avg	sd	Compound (CMP)
66_1	71.9	0.1	54.9	0.3	AcccattttcttgtagTA
67_1	77.4	1.2	75.0	0.2	TacccattttcttgtagTA
68_1	80.5	4.0	84.1	1.0	AtacccattttcttgtagTA
69_1	67.4	0.9	49.2	1.1	AtacccattttcttgTAG
70_1	77.7	0.1	67.7	2.8	CAtacccattttcttgtAG
71_1	80.1	2.6	76.3	2.5	CAtacccattttcttgTA
72_1	64.5	2.7	47.1	0.8	TGgagctaattcaggaGT
73_1	59.1	0.9	47.8	0.8	AAAtggagctaattcagGAG
74_1	59.1	1.1	43.2	2.2	TtgtccacttcttatTATT
75_1	45.0	0.9	50.2	0.8	CCccacacaatgaagcAA
76_1	52.9	1.2	40.7	0.6	TtcatcaagtcgtctgcAT
77_1	41.5	0.1	14.8	0.1	GatcttcatcaagtcgTC
78_1	68.4	2.9	54.4	0.8	AtattaacctcctatcAGT
79_1	51.4	0.8	34.7	0.3	AATattaacctcctatCAG
80_1	40.3	0.8	26.3	1.2	ATttatatcacaaagCATC
81_1	36.6	1.2	24.2	1.1	AAGTacatttatatcaCA
82_1	51.3	5.7	45.9	0.9	CAtttattgtaaagcaCAAA
83_1	56.0	2.6	30.6	1.4	ATcatttattgtaaAGCA

[0404] From these data it can be seen that single stranded LNA gapmer oligonucleotides tested in PMH generally are capable of reducing FUBP1 mRNA in hepatocytes and with a higher efficacy than observed in the Hela cells.

Example 7: The Effect of a Single Stranded LNA Antisense Oligonucleotide on HBV Parameters in HBV Infected PHH cells

[0405] Based on the hepatocyte screening in Example 6 the LNA antisense oligonucleotide with CMP ID NO: 50_1 was selected to confirm that the FUBP1 reduction also resulted in an effect on cccDNA in HBV infected PHH cells.

[0406] The HBV infected PPH cells (see Materials and Methods section "HBV infected PHH cells") were treated twice weekly for 2 weeks (i.e. the treatment was performed on day 4, 8, 11, 15, 18, 22 post infection) with 10 µM of FUBP1 LNA in PHH medium and cells were harvested at day 18 and day 25 post infection (14 and 21 days of treatment respectively) for FUBP1 mRNA knockdown and cccDNA qPCR measurement compared to the non-treated group.

[0407] The HBV cccDNA and FUBP1 mRNA reduction was measured at day 14 and day 21 following treatment. The results are shown in table 13

TABLE 13

FUBP1 mRNA and cccDNA reduction following treatment
with a single stranded LNA antisense oligonucleotide.

	FUBP1 mRNA		cccDNA	
	Mean	SD	Mean	SD
D 14				
Non treated	1.005	0.09291573	1.054	0.36328887
CMP ID NO: 50_1	0.395	0.0212132	0.337	0.02969848
D 21				
Non treated	1.016125	0.19331428	1.0275	0.28483972
CMP ID NO: 50_1	0.3975	0.02505328	0.3945	0.0205061

[0408] From this it can be seen that the oligonucleotide of CMP ID NO: 50_1 can reduce cccDNA by 60%.

[0409] In summary the examples of the present application show that a double stranded siRNA molecules, FUBP1 small molecule inhibitors, and single stranded antisense oligonucleotides targeting FUBP1 all are capable of reducing cccDNA in HBV infected cells confirming the relevance of FUBP1 in maintaining cccDNA stability.

SEQUENCE LISTING

```
Sequence total quantity: 90
SEQ ID NO: 1           moltype = DNA  length = 35056
FEATURE                Location/Qualifiers
source                 1..35056
                       mol_type = genomic DNA
                       organism = Homo sapiens
SEQUENCE: 1
gcgcaagaat gtaatagagc ttcgacggcc gccattttct ttctttctta gctgttagct    60
gagaggaagt ctctgaacag gcggcagcgg ctcttatagt gcaaccatgg cagactattc   120
aacagtgcct cccccctctt ctggctcagc tggtggcggt ggtggcggcg gtggtggtgg   180
aggagttaac gacgctttca aagatgcact gcagagagcc cggcaggtaa gtgtggaccg   240
cgcggcggaa tcccgaaagc tcacggtaat tggccgctga ctgagtaggc cgctaccctt   300
aagcgcatga ggaagaggaa agaggtgttc ttccgggctg aaatgtgaag agacacgttt   360
ccccatgttg gtaataacga ttagagacca gaacccagtt ttgtgttctt ggtgcctaat   420
ccacttagaa ccccgacgcg tgctacgcaa agaaggcctg aagtctttct cccgcttctg   480
cggcactcgc gtgtctccag tgagctagtt tagataaaga tcctcttcca ggggataaag   540
cgcagttagt ttcacacaat ttaatggaag gttctggtaa tgagtttggg aaagaactag   600
```

```
ggtctgtcct ggagccatag caagggaagg gatttatcat taaagtagcc tttacagctc    660
atttccgttc tctctcgcaa ttaaaaccgc tttcagtacc attcaccgtc acacctctac    720
aaggaaggga cttgaaagca gccttttttct gggcgggatt tacgtgtcag tctgttccac    780
cagtccgccc cccttatttc tcaaaatggc ctcaggccca ttataccaga ggtttcaatt    840
tgaatctgcc tctcagttca gagtcgtaaa ctgaccagac ctctttgtat tacgtagtgc    900
gtgcatttgc cctgaaggca ccactttccc agacgaaagc tgttaaaata gtgcgtgtat    960
tccaggaaaa taaaagatac cttaatttga actttacatt cttagatagt cccctaatat   1020
atttaacatt tcaaaatgta tggtgttggt atagatttgc atgtaagcaa aagaatccta   1080
ttctctgtga cactatgcat attgtactag gtgctgggca tttttttacta gttttaagct   1140
aatgataatt agaaaccagt gttgtgctgt gttttcgttg cattaggagt tcacttagtt   1200
aactttttac cgggacagtt gaaggaacat tgagtcaaaa ttagaattca taaaatccgt   1260
tgtaacacat ctaatgtgaa cgcattataa acatgtacct gtactttttt ataaccagaa   1320
atactaggag tagtcaacaa aaggtcatca ttaatattag ttctgcggtt ttttccacgt   1380
aatttaagaa attctgaaca tgtttagcaa caagcatata actatgacaa acactcttag   1440
cgtgttttat tagattgatt tgtaaaactt aagggaacta ttttatttac tggaaccaac   1500
tattttatttt taccagaacc agcacattgg attaattggt atcgcacact gtaggtagat   1560
actggagttt tgttttgttt tgtttgagac ggagtctccc tctgtcgccc agactggagt   1620
gcagtggtgc gatctcgtcc cactgcaacc tctgcctccc ggcttcaagc aattctcctg   1680
cctcagcctc ccgagtagct gggattacag gcgcccgcca ccatgcccgg ctaatttttt   1740
gtgttttttt tgtttgtttg tttgttttca gtagagacag ggtttcacca tgctggccaa   1800
gctggtctcg aactcctgac ctcgtgatcc gcccgtcttg gcctcccaaa gctgggatta   1860
caggcgtgag ccgctgcgcc cggccttgtt ttttttcgtt tgtttgtttt aatgcatgaa   1920
ttgtttccta ctaagaagct atgatatagt tccttgacca aatgcagatg aacaggatta   1980
tctgattaat actttaacga gagcagacaa aatatggata tttaattcat ccacttgctt   2040
tataagtgtt tatagagttt gttaggaggg gatgcccaat ttcctaagta aaggtgatat   2100
atgagcaaga cattataaaa tgaaagagag tttggccagt gaagaaaaga agcagcatgc   2160
gcagaggtgt ggggacttga gggaatggca aggacccaga agtgaagtcg ctagaagatg   2220
ggggtgggag gacgtgaaac gaggcaagga aaggtacaag cggatccaga tactggaaga   2280
ccttgtgttg caatgcttta attaaaaatt ggatttcagg ccgggtacgg tggctcacgc   2340
ctgtagtccc agcactttgg gaggctgagg caggcagatc acttgaagtc ggagtttaag   2400
accagcctgg ccaacatggt gaaaccccat ctttacaaaa aatacaaaaa gtagccgggc   2460
ttggtggcct gtgcctgtag tcctagctac ttgggaggct gaggcacaag aaccgcttca   2520
acctgggaag tggaggttgc agtgagtgga gatcacgctg ctgcactcca gcctggtcat   2580
agagcgagac cctgtctcaa aaaaaaaaaa attggatttt attattaggc taagccataa   2640
tttctttata cttcttaaaa tatagaaatt gttggcaaat tgtgaaattt attgttgtat   2700
aatgctaata gatcagttgt cctgatgtct ttgctgtaat gatttcttta taaaaatgat   2760
cttaaatctg agctccctaa ctttagtttt tgcctggaat tacccattac atttgatgat   2820
atctctaaat gtcagttgta gctgttactc tgatgatata agtggaatat acagaagcgt   2880
acttagacaa aagttaggtt aatatctgaa ctacttcctc cttgtgtatt taagagaata   2940
ttgacttaag tttctagaat cctcaactaa tcctaagttt attttctttg tctagaaac    3000
tatgctgttt ttgtttttgg aaggaagaga tataggcata gtttcctgct ctcaaggagc   3060
ttcaaaggct gtaccagtgg ggatgccatt ggtattttta gctggatagt tgttattcag   3120
aaaagcagga caagtaatta tgattcctgg tccgtacctg gtaatgccag taatgttaac   3180
tctagctggt tgttgacatc tggtcattta gttgccaatc ttctttttt ttttctggct    3240
tttatgtgaa attttttagat ttttataata tcctgagcta aattcaacac agggacacca   3300
gattgctgct ttagttcagg gtttccagcc tgtgcactta agaaatttat ttttgtgtat   3360
atcaagctgt aactccagag attgggattg tttgattggg tctttagcag tggtactaat   3420
agcaacttct gtctctagaa cattggaaaa ttaaaatgtg tttatctacc gttttttttcc   3480
tcgaggttat atgaaggtag aaatgaatca gactagatga ttagctaagc gagactatta   3540
accctcatcc cttcccctct agacaactat gaaattagtc attatgtatt cgatccttct   3600
tgcagtctct tctctgacag ttataaaagt gatttaggct gcataatgtt gtttgaatga   3660
aatgaaaata tagactagag ctgttttttt ttttatttcc atcagtctct tcagtgaaaa   3720
ctaacatttg agcatgattc ttttttttaaa tcatttgtg acagtttagc aaggcttgtg   3780
ataagcaagt tatggtatgg taatatttct agtgtccacg tttcttcaca tgtctggtgt   3840
atgggaacta ctaactccat caggaccttg cctatagtag gtactcaaca tttactgaat   3900
taaatcaata aacatttttta atgaattaca gtacaagtca gacctctgta tctgtgggct   3960
ctgcatctgc aaaattcagcc aaccatggat cagaaatatt agaaaaatgg aagaacagtc   4020
cagcaataca agtaatatga ataaaaacaa tacaacaact atgtacattg tatcaggtat   4080
tataagtaat ttagagatgc tttaagtata ctgaaggatt tgcgtaggtt atatgcagat   4140
actgtaccat tttatataag gaacttgagc atctgtggat tttggtattt gcatggttca   4200
tggaaccaat ccccaggga tactgaggga ctatagttga tcataccacc tgattttaga    4260
gattttctga gtctcagaag ttaattaagt aaactacaat agtctgttct taacctcgga   4320
ggatacattc caagaacctc agtgaatatc tgaaccacac gatagtattg aatccaaatt   4380
atacacggta atatttttttc ctatacatat gtatctataa agtttaaatt ctaaatcaga   4440
cacagtatta acgataataa taaattagtg caagactggg catagtggtt cacacctata   4500
atcttaacac tttaactatg acgttgtctt tgaaaagaaa tcagctagcc aaggtggctc   4560
atggctgtaa tcctagtgtt ttgagaagct aagtcaggaa gattgcttga gcccaggagt   4620
ctgagaccac cctaggcaac atggtgaaac cctgtttcta taaaaaatac caaaaaaatag   4680
gctgggcgcg gtggctcacg cctgtaatct cagcactttg ggaggctgag gcgggtggat   4740
cacgaggtca ggagattgag accgtcctgg ctaacacggt gaaaccccgt ctctactaaa   4800
aatacaaaaa attagccggg cgtagtgggc gcctgtagtc ccagctactc aggagactga   4860
ggcaggagaa tggcgtgaac ccgggaggtg gagcttgcag tgaaccgaga tcgcgccact   4920
gcatgccagc ctgggcaaca gagcgagact ccatctcaaa aaaaaaaaaa aaaaaaccaa   4980
aaaattagcc agacgtggtg gtgcttgcct gtagtaccag ctatccagaa ggctgaggtg   5040
ggaggattgc ttgaacttgg gaggtcaagt ctagaatgtt gacaatgttg ggtcctttat   5100
gtagttgcat aagtgagcca tgatcgtgcc actgcactac atccttgggc aacagcctga   5160
```

-continued

```
ccctgtctca aaattttaat ttaattaaaa aaataaaata gaacaattac aacaatacac  5220
tgtattactg gacaagaagg gcaaatttaa aaaaaattaa accaatatgc tataataagt  5280
tatatgaatg aggggaccct cccctacccc agaatatctg attgtactat atcataggta  5340
actgaaaccg tgaagagcaa aaactgaaga taaagagact actgtgtctt ttaagtttct  5400
tttcaactcc caaattcttg gatttctcac ctcttggctt cctcaataga ggtgagaaat  5460
gttaaagtag tgaaaacagg aaaaataact tactcattca agaagtagat aatggtccag  5520
atggaaagct tgaattattt ttgtaaaact aaaaattaaat aaagtagcca ggcatggtgg  5580
cttacgcctg taatcccagc actttgggag gctgaggcgg gtggatcact tgcggtcagg  5640
agttcaagac cagcctggcc aacaaggtga aaccctgtct atactaaaaa tacaaaaatt  5700
agctgagcat ggtggcgggc gcctgtaatc ccagctactc gagaggctga ggcaggagaa  5760
tcgcttgaac ttggggggcg gacattgcag tgagcccaga tcacgccact gcactctagt  5820
ctgggtaaca tcttgagact ccatctcaaa taataataat aattaaataa agtaaaaagt  5880
ttcccacacc tcataaatgt ctaataaaaa ttgaatatgt tgagttcaag tactctgaaa  5940
aaggagttga atatagttgg aggttggttt ttaggaatta ctattttct taaattaact  6000
atccttgtag tcacctagga attgtgtatt ttctatagat cttagaaaat tatcaaatct  6060
acagttcatt ttgttttttc agttttttt ttttttttaa gagatggagt cttgctgtat  6120
tagcgttgaa ctcctggcct cagccagttc tcccatctca gcttctgaag tagctggggc  6180
tgcaggtgcc actgagcctg gcttcttat tggtattttt attaaacact tttctctaat  6240
gtctttgtaa cagttctcag tttttgaaat gctgttactg tttctttagt gtgaactgtc  6300
aactttcatt ttttcttttc ttttcttttc ttttcttttt ttctttttg agacagagtc  6360
tcgctctgtc acccagactg gagtgcagtg gtgcgatctt ggctcactgc gacctctgcc  6420
tcccgggttc aagtgattct cctgcctgag cctcccagt agctggaatt acaggtgcgc  6480
accactgtgc ctggctaatt ttttttttt ttttttgtat tttagtaga ggtggtgttt  6540
caccatgtca gtcaggctgg tcttgaactc ctgaactcat gatccccccc gccctgcctc  6600
ggcctcccaa agtgctggga ttacaggcat gagccaccac gcctggcctc agctttcatt  6660
ttcatttggt tagtttttga actattcagt gggtaaagtt gtataaataa gtgtcttttc  6720
tctgtataga agtgtcttgg agttcaagga gtgctgcttt gcaaactcat agagtattta  6780
taaaagctaa ctgcagaagg tattcatagg ctaaaccgtt tcctattctt ggtagcacca  6840
ttttctctgg cctgaaatac tttccttcta ctattagtgc ctgtcgatac ccagcagtgt  6900
atttactttc ctgaggaaca attcaaatgc taagtgcttt aagacctaag ggtggaaaag  6960
cagtgttttc aggcattatt aggaaaataa gatttaaatt agacacccag aaacaaaaac  7020
aggtttgtaa ttggtaaagt gaaagatggt taaagaaggt tagattgacc aaagcgagaa  7080
tttacctttt ttttttttt tttttgaga cagagtctca tgccgttgcc caggctggag  7140
tgcagtggcg tgatcttggc tcaccgtaac ctccacttcc tgggttcaag cagttctccc  7200
acctgagcct cccgagtagc tgggtgacat gcgccaccac gctcagctaa tttctttgta  7260
ttttagtag agactgggtt tctccatgtt ggtcaggctg gtctcgaact cctgacctca  7320
gtgatctgcc cgccttggcc tcccaaaatg ctgggattac aggcatgagc cactgtgtcc  7380
ggctgagagt gtaccttttt ttttttttat caagcaatct agtacttgat cctaataatc  7440
tttgtggtag gtgtttgcat ttttagatga ggaaaaggga aatctatgag tcctaggaaa  7500
tacagttggt atatgggaac tgatatgtaa ttagacttaa gtgatccatg ttgaatttat  7560
gacttaagca cttaactata atcttaacct ctccagttgt ctgatgaagt tagtatatgg  7620
gaactgatac atagacttaa gtgatccatg ttgaatttat gacttcagca cttaactata  7680
attttaccct ctccagttgt ctgatgataa taaaaacttg aagcagttat ccatatgggg  7740
atctctttgg ggaatcccag tcaccaaaag ttaggttttc ttaatattt tttcatggaa  7800
gatttcaaat atactcaaaa ttgaaagaat tatataataa attctcatga gcccatcaca  7860
catcaataat gaatgtacag cattgcagtg tggagcttgg cctattgctg accactcagc  7920
aatgtggcag aaccactcca tgattcccca tggaaatggg aactacttcg gttgtccttt  7980
tatagaaaaa ttcagtaagt atctgctgat tgtgccctac ttgtgacttg aagccaggtt  8040
tttttttttg ttttttattt ttttgtttt gtttttgtttg taacagtctt gctctgtcat  8100
ccaagagggg catgatattg gtgcactgca acctccacct cctgggttca agtgattctc  8160
gtgcctcagc ctcccgagta gctgggacta tgggcgtgca ccaccacacc tggctaattt  8220
ttgtatttag tagagatgga gtttcatcat gttgcccagg ctgctctcga actcctgagc  8280
tcaagcaatc tacccacctc caccctccaa agtgctaaga ttacaggcat gagccaccat  8340
gacagcaaag ctgggtattt cttaaattgg ttcagtcagg tgcaataaat tatttgccct  8400
actctaaaat ttaaaaatct tctaagaatt atggttttgc agctgaggtt ttttttaagac  8460
tcgagctccc tggactccta catatatcct tagaacaaca ttgtccaata gaagtacaat  8520
gtgagccaca tgttttgttt aacccagcat atccaaaata ttaccccctt tgcatgtgct  8580
taatataaaa atttaagatg atttatattc caggtttca atattcagtg agtaatttta  8640
cacttagagc aagtatcatt tcagactagt cacattttga gtactcaata accacatatg  8700
gctagtggct accttactag atagcatagc ctttgagtcc cacaaagtgt tcacattcta  8760
tgtgtttaca cacatctttt gaagtgtcat gacagagcca atgagatcc agattttctt  8820
taaatctggt ctctcttccg gattcctagt tgattacttc tttgttgctt cttataggga  8880
tgtatcagaa gtttaataat cttatgatta ttatgttaac tgcctgaagt attaatggta  8940
gcaataaatt gaactataat tttaattttt caagtaaatt ttcttatgtg atactaaata  9000
tttcacatat acatgtgatca ggaagtaaat gggggtaatt tagtaatgaa tagtatataa  9060
tagtttatga tggtattttt cttttttttt tttagattgc agcaaaaatt ggaggtgatg  9120
cagggacatc actgaattca aatgactatg gttatgggg acaaaaaaga cctttagaag  9180
atggaggtaa gttatactct aagtattta aattgttttt cagagtgttt agttgaagtg  9240
attctcggta tttttctgtt attttattga gatattaact tttattataa ggttgttaaa  9300
attgtaagct gtatattggc ctaaaagggg ggaaagaaaa ctagcaagg taagtaaaat  9360
ttgaaagaaa tttttaaaa aatttttaa aaaaggaagt tttgtcttaa tagaaaacaa  9420
tttattttcc ctcttttagg attgtgccag attgaaagtt tgcacagacg tcttgttaat  9480
aatattaaaa aacatataaa ttgcttagaa gacacttcac tggttttact catacgtgaa  9540
tggatttaa tatgctgtat tttcgtcatt tttctatttc caattgcacc ttaaaggttg  9600
aaattccttat agtttgctac tctagtgtgt tgcaggttat accatttttt ttttaatgtt  9660
ctttactttc agtactttttg tgtttcacgt ttagctttaa acctgtggat taaaacagtg  9720
```

-continued

```
gatttacagt gctatgtatt tttaaaaatc gtaatcgttg aagcttctga acttagaagt   9780
ctgcatgtat tttttgtttt aggtttgata tatgagtttt gatacatttt cttttttaccc   9840
ttttttttaa agggaggatt ctcactgagg ctatagaatg tatttgtagc ttttgaccag   9900
gagaacttgg tttcctttta tttaagtgtc tttcatattt atagtgaggt ttttaatgta   9960
gaaaaaaaat gagctaatga tgcttcagat gttgtatgta atgtattctt tttatttttat  10020
gtgtagatgg ctcttggaca agtccgagca gtacaacaca ctgggaggga atgccctctc  10080
cttttaaagg caggaatttt tatttattac ctgtgttcag tatgtaaacg tgaaataaac  10140
cagtggattc ttaaatggac acaaatattt cttggattat gtgtctgcgc atattttatt  10200
tttgctgcac aacattctga tgtttaatca tttaagtttg aaggggggag gagaatgtag  10260
tactttgagc tataggttgt ctgttccaag gtatgcattg tattcatctg tgtaatggat  10320
ttaggtgaag gtagtcatgt agttgctttg agattttatt ttttgctaa agttttatgc  10380
agtgaaatgt ttgtttataa ataatagaac agtttaggtg gagattgcct tgtaatgttg  10440
tgggggggttt ttttgttttt tttttttttgg gcatagcttt gtggtcactg tcagatacac  10500
tttaatatgt cagattttttg tagtttgata ggcttttctc cccccagtct tcagttcctg  10560
aggtggaagc atcattagcc tttagcatgt gatattttgc tagtaatgga cctaaagtac  10620
gttgtcttgt gtcattctaa tgtgcttaac atacattaag gtcagtgatt tcttaagaat  10680
cagataacta ttttaatgtc tgtgcatctt ttgaacgtga agagaatgaa gtatcgtttc  10740
tttttttagat tactgagttg gggttgaatt ttggcagttg tggttcaaat gataaaccat  10800
accttcagat atttcaataa atgtttatat tgttatttat tcttgttttgg gagggggagaa  10860
gctttgtact taattggcaa aaaattaaaa gacacttaat tttgcagatc aaccagatgc  10920
taagaaagtt gctcctcaaa atgactgtaa gtattccttt taaactgggt caaaagctaa  10980
agtaacaatt tcaatgttaa gattttgttc atattattgg gtatctttga agttagttgg  11040
tttatgagta tttttggatca gctagcctga atttctttgt aaatatatac cttttttctcc  11100
tattttacaa tctgtcccat taattgtggc cgggtatatg taaagattgg ttgctgaatt  11160
attttacata tttaacaact ccaattcttg atctatactt gtacaacttg aaaaaggaaa  11220
ttattttgtt ctgtgccatt gctaatataa tgtcttccct ttcattggct ctcttgcccc  11280
tgtcaatgcc aatataaata ttgtgtaaaa aatttgactc tcttcaaggt gttgtctgta  11340
cattagggag agattttttga atgtttgatg tgattgtgtt gttaaatata taagtaaatt  11400
taaattttga attttttgttt tattttttatt taatttttttg atagcttttg gaacacagtt  11460
accaccgatg catcagcagc aaaggtatag tcacaagatt ttcaaaaagt actctgcaag  11520
ttttggttga gctgtatgta aaaacacaac cacattggtg tatattgaat gtgtgtctgt  11580
gtattttttg gtgtacctag ttcatatcac tccccttggg aaggtaccat aaagtgatga  11640
ttttttcttt gagtgagaaa aatttgtgat ttggagagat aggtggaatt aaccacattt  11700
tagaagaaca ggggtgaatt agagtaactg ttaagatgac attctctaaa ctccacttca  11760
acttctttac agttaatgcc ttcagactgt tccattcatc atcccttctt cacttgatgt  11820
gtcatcttaa atttcttaat ttaactactc aagtaataag atcatatttt ttgacatgac  11880
tctgagccta gaaccttagt ttaagccatt gggagacatt agacttccat ttttattaat  11940
agattatctt ttatttgtaa acaaagtatc tttcattgaa ggaaaatggt gctttctgtt  12000
atttcttagc agatctgtaa tgacagaaga atacaaagtt ccagatggaa tggttggatt  12060
cagtaagtaa cttgattttt aaagtttttga aaacatgact aaaacatact ttagaatctt  12120
tcaaccaaaa aaaaaattttt ttttttttcta actagtaatt ggcagaggag gtgaacagat  12180
ctcacgcata caacaggaat ctggatgcaa aatacagata gctcctggta atgttacatt  12240
ctcatggtat tttcagtgtg actagaaaac tagcttttttt ttttttttaa gccttctagt  12300
aacaataatg ctactttttaa tcttttgacc tgaagttatc tgtttgtttt aaattgtaga  12360
cagtggtggc cttccagaaa ggtcctgtat gttaactgga acacctgaat ctgtccagta  12420
agtttgaaaa atcttaaaaa tctacttaag taacaacagc agaactcttt gaattttgtc  12480
tcttctcttt gttactgctt tattttacac tgtggtttcg ctgccacctt ccctcaaagt  12540
cctccaactc ctttgaagtt tatgcctcat gcctttctca ggtggggttc atcatctgaa  12600
tcattaaaca cagaaaatgg ttaaaacaac tccatatcta ctccagtctc tacttgtaaa  12660
gccacgtgta gcctggagaa gaatgcacag tcaggtcgac tggtgacact taaaactcag  12720
acattaagct caagtggact gttgtgttgc ctgcatttcc ctagttccat tcacttttcc  12780
actcctctcc caggctcttt aatactgtat ttccccacct ccaaatcttc agcatctaac  12840
cccacgctct cccacttaag cttatttact gagaaaatgg aagcaaatga taagaagctt  12900
ttcttttttcc ctaccactaa acctaccagc cttcatttttt cctctgttca catagtactc  12960
aggtaattgc tttcctttttg tgttcgagtg cctaaagcca gccctttctt cctactgaag  13020
gttcagcttg cagttgtact ttcttctgca ttgttagttc tccctcatta ctaggttttt  13080
tcttttctttt tttttgagac gaagtctcgc tctgttgcca ggctggagtg caggggcacg  13140
atctcggctc actgcaacct ccgcctcccg agtagctggg actaccggtg catgccacca  13200
cacccagcta attttttgtat tttttttagta gagacagggt ttcaccatgt tggccgggat  13260
ggtctcaatc tcttgacctc gtgatccacc cacctcggcc tcccaaagtc ctgggattac  13320
aggcgtgagc cacttccccc agcctgattt tctttatggc actttccaaa taatgtgtta  13380
ttcatttgac atgttatttt tacttattta atgaaatgaa gccactctag caggaaccct  13440
gtttctttga gtgctattac cccattacct agaatagcac ctgcacatag ttgatatttta  13500
aatatttgtt gaatgaataa ttgtagcata tgagtaagca aaatggtagt ttaaaaatgt  13560
aaataaatca tttagttctt ggaagaatca gtttaattct gagataactt tagcattaga  13620
gttctttctt ggaaattttg gactattctt aaaaataaaa attgtatatc tagaaaattt  13680
ttttgcataa tctctcaatc tttgacccttt gatggcattt tctttcagtt aaaagtaaaa  13740
gcattgttaa agttagcatc aaggcaccta atcctgaact gggataggag gagtacttgg  13800
ttatattgtt ttatatttct ctatttgaat aagcttgggt atgctacagc ttactattta  13860
aatattaatt tgttaacagg tcagcaaaac ggttactgga ccagattgtt gaaaaaggaa  13920
gaccagctcc tggcttccat catggcgatg gaccgggaaa tgcagttcaa gaaatcatga  13980
ttccagctag caaggcagga ttagtcattg gaaaaggggg agaaactatt aaacagcttc  14040
aggtattgtt attttttgtga aatggctact tttgatctgt tttgatgccc attttttgtcc  14100
acttccttttt gttaatatat attatttcta tgattgtaac aggaacgggc tggagttaaa  14160
atggttatga ttcaagacgg gccgcagaac actggtgctg acaaacctct taggattaca  14220
ggagaccat ataaagttca agtaaactta actttatact ttataaagaa agagtgggtt  14280
```

-continued

```
gaatggggtt gggcaaaata tgcatgaata attaaaatgt tttgagacat gctttctaaa   14340
ttagctaact ttttctgctt tagcaagcca aggaaatggt gttagagtta attcgtgatc   14400
aaggcggttt cagagaagtt cggaatgagt atgggtcaag aataggagga aatgaaggga   14460
tagatgtaag taaaaatacc cattcagaaa tggttgtatg ctaattcata aatataaatag   14520
tgttttctgt tttgtgttaa gtagctctaa cattgttatc cttttatttc accttttatac   14580
tttagaatac agaattctat atatcttgtt accctatta ctataaatat agaattatat   14640
gtacttttat gatttgaggc agattttcag gaaatggcgc ttttttaaaa tacttttttt   14700
tactttaaac cctgagaagc tagctttctt aatacttagt ctttttaca taaggtcccc   14760
attccaagat ttgctgttgg cattgtaata ggaagaaatg gagagatgat caaaaaaata   14820
caaaatgatg ctggtgttcg cattcagttt aagccaggtg agtacatata ataatcttgt   14880
aagtgttggc agcagtgagt tttgacatac atttattgtt taattaattt tgtttctttg   14940
ttttgaagat gatgggacaa cacccgaaag gatagcacaa ataacaggac ctccagaccg   15000
atgtcaacat gctgcagaaa ttattacaga ccttcttcga agtgttcagg tttgatagaa   15060
agttaacatt ttcattttt gttttatgg aaaagtattt tccttcatga aatctgaagt   15120
tacctctata tcagagtctg cttgatgatg ctttattaat ggagaaagtt taaattgctt   15180
taaggtaaag atcttggagc aggaagaact actaccttaa gtgctacctt atttactctt   15240
gtattttaaa aaatgttatt acttattagg gccagttcat ctctacattt ctgattcagg   15300
tatatgagag ctgggaaaaa taaacttaat aattttatca tgaaacaaaa gtatttctgt   15360
gctgactctt cgttcttgtc tttccctctc tataggctgg taatcctggt ggacctggac   15420
ctggtggtcg aggaagaggt agaggtcaag gcaactggaa catgggacca cctggtggac   15480
tacaggaatt taattttatt gtgccaactg ggaaaactgg attaataata ggaaaaggca   15540
atgtatttta aactcttaat gttttaacac attattcatt tttctggaca ctttctgttg   15600
ctgtcgtaaa caagtggcaa tgctttttct ctgaccgtat tttagtagaa aagaattctt   15660
atgttaatat gtaacaagta aaacataaat gagggatctc atgtatattt agagaaaagg   15720
caggatttta atcttactag cttctagaga aagcgaacta agagataatt attagcataa   15780
gaaatgtctt ttgacccaaa aagtggtttg agtgttttg tttgtgcatt ttggtttttc   15840
ccgactcata ttttaaaaat ttgaatgttt ataagtgtat tagtttatat ttacactgct   15900
tttaaaagca gttaattcaa atattttatt ataatcacat taagtttatg tttaaacata   15960
ctaagtaaat gtaaatgtat tttaagagaa gcatgaaatg cttcctaaaa tttgatttc   16020
agtgtagaat attaaatgaa aaatcttaat acaatattgt caattaggat actgaccaaa   16080
ccatatttt aatggcccat ttaattgtga ccattttctt ctaaatagct cctagtacac   16140
ccttgaaacc tttagagaaa ttactgtctt ttgatttag gaggtgaaac cataaaaagc   16200
ataagccagc agtctggtgc aagaatagaa cttcagagaa atcctccacc aaatgcagat   16260
cctaatatga agttatttac aattcgtggc actccacaac agatagacta tgctcggcaa   16320
ctcatagaag aaaagattgg tgtgagtata ctttaaactt ttaatttta gtgtagaccc   16380
ttagactgta gttaaattaa gacgtttatt caaatacatc aaaggaaaat gtatcattac   16440
tagtcagcat ttatagattt catgatatgt ataatagata caacgtgaag attttccagc   16500
aatgaaaata acctaattaa atgtgtagtt acaggttttg agaacaacct tacatttggg   16560
tgtggctaga taagaggagg gtagtgttac ctgtaggcat gatattagtg ttggtgtagg   16620
attgtggaac atacttgaag gacatagtta acgggaattc attatttatt aagattttac   16680
tctactgaac cccagcgagg caaacaagat aaatcagata catctgccac tctacagtag   16740
aaattcataa atcctaggtt ttggactggc tcacagatca tctggggggta ttaaaatgta   16800
gactatttgg gtgcttccca ctcctacaca taatatacag agacagacct gactcaaaat   16860
gtctgggata gggcccagca tctaattta catgatgtt tgggtgattc tggtgcacag   16920
acaaatttgg aaattttcc tccaggaaag ttttctgtta gaacaaaaaa gtatgaaacg   16980
ctttgactgc ttttttgtaa gtgaggcaga cagtgtctta ctggagtttt taacacaaag   17040
tgtgcagggg gcatcttaaa ttattagaat tgcctaggaa aaattatttt tggtgtttgc   17100
catgctgtga atggggtgcg tgcaaaacaa ggttgacact gtttctgtca acttgggagg   17160
acaaaaatag atgtacaaaa gacgcttaag gaacgtctta caaaatgtac acaaacattg   17220
tgagttctgt tgtatggata gtttaatggt tcaaaaaatg aagaggatgt ttgtagagta   17280
atgatagagg ttgatgccat ggcaaaaaaa tgaatagcac tcatcttggt tttttatttt   17340
acagtattag cctaacatgc atgcgtcagg atttctgttg gattcatgga aaaacaagat   17400
agattgtttt ggagaagcaa tttggtgtgg tgattaagac catggactct gtagtcgggt   17460
tgcctcagtt cagaattctg tgaatcattt actggttctg tgaccttgga gaagttactc   17520
aggctttct gtgccttggt ttcctcctgt aaaatgaaga taatggtatc tacttcatag   17580
agttgtaggg attaaatgaa tgttcatgtg tgtgtcactt aaaagaatgc ctgccacata   17640
accctaaaaa atgttgctac tttttcagta ttatttttac tacttggaaa gaataggtca   17700
tggatagtaa gtgagagatg acagcaattg gagatttaaa gagacaagtt aataaaaaga   17760
actttaaagg tcatgaagtt tagtttccct attttgcagt gagaaattta caacaaaatg   17820
tagtgttagc attttgtcca acatgtcgcc tggttgtgtt aactactcag aaggagcatt   17880
ttaggacagt taagtgtaat gtctttgttg gcttaaccaa acagaaatca gttaagcatt   17940
attaatgatg tggcatgcat gcatgataag gatataaaat atcctgattt attgaaggaa   18000
ttaaaaaggg aattttttgtg ctattaaaca tcatgataca tgaaaggcca aaaaggatat   18060
aaattattga tctgaatggg attttagtga cagaaatagg ttgtgaggtg gattttagtt   18120
tcatagtgaa acaactagct attaacgtta tcagtgaaat gttcagaaga cgatgataca   18180
ggacccaagc ggtttgggaa tatattgtca agacggttag tccatgttag gcaagtagat   18240
atagagagaa gagctgaaga taaggacctg atttctgtga gtaagatgga aaaagatagt   18300
aaaagtaaga atgataaaag gaaagggaac tactgctgtc cgcaaacaag ctaaaggaaa   18360
tttaattgct aatttaaatt taatttaatt taaattgcat ttaattgcta ctgctattga   18420
ttttagtgaa ttttacatgt ctcattatta tggcagatga aatagttttt gcaaaatgaa   18480
tgaggaaagg aaggaaaacc taaaatttgt tatttgtgat tataagaggg tagaaatgga   18540
tgattatttg gccgtagagt tgtttaacca ttactgtgtt tttctgattt ttctatgtca   18600
tcctttttt tgtagggccc agtaaatcct ttagggccac ctgtacccca tgggcccat   18660
ggtgtcccag gccccatgg acctcctggg cctccaggc ctggaactcc aatgggacca   18720
tacaaccctg caccttataa tcctggacca ccaggcccgg ctcctcagta agtattgggt   18780
ttagttctgg gctttcccca aagattctag ttttgggact gttttttatg ctgattttc   18840
```

-continued

```
ttttcagtgg tcctccagcc ccatatgctc cccagggatg gggaaatgca tatccacact  18900
ggcagcagca ggctcctcct gatccaggta gaagatgctt attatttgtg tgttatctgt  18960
attattttcc actcctgtta cattattaaa ttttctagtg ttgattctac atttgtatgc  19020
atcaccttca ctcactttac tctttcaaca gtgttaggca ctgcctctac cccagtgtat  19080
aggactgaca tgaatatgag ctctgctttt atggaatttc tttctacttg cctttggctt  19140
atgagttgat acagtagaat gataaaaagct aaaagctgca ggaaagagca cagtgtcata  19200
ggtttttggat accagtgctg tcaaatgtgt agtatgttca ttgtgacatt atctgtggaa  19260
aaatagtttt tacttattta gaaaagtatg tgatagggggc tgggcactga ggctcacgtc  19320
tgtaatccca cctattcaga agtctgaggc tgaggcttga gaccagcctg ggcaatatgg  19380
caagacccat ctctaaaaaa atttgtttta aattagccgg gcagggtggt atgtgcctgt  19440
acttgaggct gaggtgggag gattgcttga gcctagaagt tctagattgt attgagctgt  19500
gggcacacca ttgcactccc ctgggcaaca gaatgggatc ccatctctta aaaaattatg  19560
tatatatgta acagtctata taaatatata tataacag tccaacagag tgttaagtat  19620
tggggcatta aacttccaaa ttgtcaaata agatatctgt tctagtactc ctcatatgac  19680
aacttcatgt gagtaaaaat caggcctgta tttaataact gcatgctaaa gcccaaatac  19740
gtttaattat tttctatatt cagatttatt tttatcctat tatattctgg catctttcca  19800
taccctgtag tttgctttcc atcttggtca aagagtcatt ctttgaaacc aattaacatt  19860
tatttatgat cctttttttct cacctgctac ctaccttta ggcctcttc actcgtgtag  19920
tttcaaggaa aatatactca attatggaat actttctaca cataatacat ttatcccaaa  19980
aaaacttgaa gtaatttatg taaatgaaat tgcttgatta acttcatagg aagtgtgctg  20040
tttggaatat gatgacacag catgacagtt accaagcatg acttgagttg tgctaggata  20100
taggtgtgca aagttggggt tgtgttctat agcaaaagaa tgcactccca gcgtaagcaa  20160
cactgatgga aggggctcac agggtacagg atataatact cttaacaact aattttttgga  20220
gaatgaaagg gcttttcttt ccctcttgtt ggctgattgg gatggtataa ttaatgggat  20280
tgaagagttt gagtaggtta aaaggcagat tcatattggg taacttggat ctgccaggat  20340
tgtatttttg agcactactg ggtggttagc atgattgagg aaaaatgatg ggaataagaa  20400
gtggaagtgg tctttgtatc acaagttgaa tttctcactt ggagtagtag tgaagtcact  20460
actgtaagag ctggtcagtg aatgtggttg cagcatggcc tttgggcaag aagtaaccca  20520
tttaactaaa accagctggt tggccccact cagatttatc aaagggttac tgggtccctg  20580
ggggtggata ttgcttatat tagacttaga atagcatact gttttaatat tatatgaact  20640
aaaatgtttc tttaaaaaaa gagtggtctg ttaatgaatt tatgtagtgg tcaagaattt  20700
agacttcaga gtcaaataaa cctatatcag tcctagtcct acagtttact aattgtgaga  20760
tgtcaagcaa gttttttgaac tcctctaagc ctctgttttc ttatctataa attaataaat  20820
gaatgaatcg ggttgagtga atatttagta aattcttagt acatactagt tatttgtaac  20880
tgtgagactg gtttttttggt atggttttca catttgggag tagaaatacc acttcctaaa  20940
gtctgtttta tctcaaattc tctatccagg catagtgtaa agtgaaatac ctagatttct  21000
tgattaaatat acagataatg gccagacgcc atggctaaaa cctgtgacgc tagcacttcg  21060
gaaggctgag gcgggcggat cacttgaggt caggagttgg agaccagcct ggcaaacatg  21120
gcgaaaccct gtctctacta aaaatacaaa aattagctgg atgtggtggc aggtgtctgt  21180
aatcccagct acttaggagg ctgagacagg agaatccatt ccactgccct  21240
ccagcctggg caacagagtg agacacttca tctcaaaaaa aaaaaaaaaa tacagataat  21300
gacactattg agatatgtaa acatccaaca taaaaaagca gtacattggg caattgagaa  21360
aagtttggca ggtgtcctaa taacaccttа caattatatt accttgcaat tttcttttttc  21420
tagaaattct tgctcatttt tctcaaactt gactaacctt tattaagcag ctgagaattg  21480
ctactgttca gaatgaaggt ataatagaaa atttaaaagt tttaattgta tggtattcct  21540
agtataaaag acaatcaagt tttttctggt ttctagaaga ttgaaaggtc tcaaactgtt  21600
cgcacttcag ttgatgtggg agatgagtga gggtcagtca agtgtagagg agacaacata  21660
gctaaaagcc gagacacggg catagtgatt tctgaaaagt acaagcactg tgttgtggat  21720
ggagcttggg tgttgaagag atacaaccaa agggagagat gaggctgaaa tggaaaggat  21780
aagccagggg cttcaggctg ttaaagatgt tgaatgatag gctgtactta agctttgaac  21840
atcctgaaga ttgcagagaa tgaagaattt caagtaggat gacatcaagt ttataacatt  21900
caatgtgaaa taggatgaaa tgtggctaat cttttttaag atttttatat tttctcttca  21960
ttgaaaataa ggacaaagtt cattgttcta aaataatttt ctttcttata caggttagtc  22020
cctagaaatg tttttttttа gtcatcttct agatagagct gtttgtgctt gaggcgaaac  22080
caatttagaa aaaaacaagg gcacaggatg gtttgagaca gagcattgga tttggagcca  22140
aaagaccttt attcatatcc cagttatgcc actcagtagt ttttaaccta ggaaagtcac  22200
ttagtctctg agacttggtt ttctttaata aaagctagtg gtgacaacaa aaataataga  22260
aaatttaagt tttcatatac ttttttttaat gttctagtta attttggggg ctgtatattt  22320
gccagagagc tgggcggggg ggtggtgtgt gtgtgtgtga gagagaatta cagacttcac  22380
atgcaaacct tgaactttca tttattttag ctaaggcagg aacggatcca aattcagcag  22440
cttgggctgc ttattacgct cactattatc aacagcaagg acagccacca ccagcagccc  22500
ctgcaggtgc accaactaca actcaaacta atggacaagg taactacaga acttattgta  22560
tgtgaaagcc aaaagttgtg cttggaatta tatatgaagt acatcactgt ataatacc ta  22620
aaatttctga cattattttaa ttataatttа agcagacttt tccttttttta aattgttact  22680
ttgaaaatagt ttcaaacttt cataaaagtt gtaagaagaa tacccggagc acccagatag  22740
cccagctgtt aatgtttttat tcccttgttt actctgttac catctatatc atacatgtgt  22800
atctgtattt attttaaata tccattataa ttagttctga aggaagctaa ttatatccag  22860
cataattagt tctgaaagaa atcactgaag actaaactgt agacatgatg tcctgttggg  22920
tacttcttca aaaatacata accatcatac agccctccaa atcagtaagt caacattgat  22980
atattaagtt ctttatatca ttatggatтt ctgattttttg cagtaggtta tattttttaaa  23040
gtattttat aatttttgata atcagattat cctggatttg gccgtcagg gagtatattc  23100
agggtggtgc ctatatcctt ttgaaatacc tgtcattctt ttaagcactt ccatactgtc  23160
tggcacagta agattgttat tttgtgcttt ctgggctcca gccctcgagt cagctatttc  23220
ttcaaggagc tcttattcct tttagtagag tatggtagtt agaaacgaga cttgagtatg  23280
cttgttgcta ctgaggtgta attgcttcta gcttctttca acaggcagaa ctaggaaata  23340
tatttacata catgcatacc tacatacaca caccaaaaaa cacataaaca tcgatatatg  23400
```

-continued

```
tatatatttt taaaaactat gttcatatca ccacatccgt ttcagcattt tggggttttc    23460
aagccttttc ctttttttgta cttgtttaca aacgtgagaa acctggtgtc ctcagtgtat    23520
ttccttattt gacatgcatt tacttatatg ttcaacataa ccagtcttca aacaggttgg    23580
ttttcttttc tgtccaccac ctctgtaccc ccagtacctt ctatctttgg cactgttagg    23640
gatgccacca ccacatagta cttccctcct accctcactg tcacattgca ggccctgcc    23700
agctcctgca cccaaggaaa cggccaaaat tgccttttaa aaactttaat tctgttttt    23760
gttttgtttt gcttgaacaa ggagatcagc agaatccagc cccagctgga caggttgatt    23820
ataccaaggc ttgggaagag tactacaaga aaatgggtat gttttataca tttcttgaaa    23880
atacatactt aattaaattg aaacaaatta ttctcttcag gaagagaata attgaataaa    23940
atcactggac ttgtaaacat atcaagacag ttgtaaatta tagttttaa atttgtggtt    24000
atatggcaag gaaattttt tttctaattg catttgtcaa ccagttatta attgaaacta    24060
gaaatgtcct tactggtaca tacaatatta acattactat acttttgaca atgacagtta    24120
tatatattat cagtctaaat gacataaggt taaattttaa tgtgtcaggc gaaaattgtg    24180
tgtgatacca ttatttttgc tgcaagataa gcaggtaaga agtaatctgc agtgacggaa    24240
agtaaccaag tgatggaacc agaatctggc ttccaagagg gtctgagtcc caagcttgtc    24300
tcccaaattt gtctctttag gaaccatttg gaacctgata ctatacttct ggacaaatca    24360
ctatatttca gctgcttttg ctcttagtca tttaaaatta ttacatacca cagctagatg    24420
tcacaaatga aagctaaatt ggtaagcttg gttatccttc actagcagaa aaagaaccta    24480
taggtggtag agttttgtca taagagaacg gtctacttgg gattttcaat gttttctttg    24540
ccaaagaatg tttctcattc tccacagaaa gaaaaatttc cagaaaggtg atgattttaa    24600
tcttctagat gtaaaattac atatacctga tgataaagtt gttttgcaca actggtttct    24660
ttttaaagaa aaattgtttt tcctcttaga atggcttcct aggagagtca tgttccgtct    24720
ctttctgagg ctttaacaga ttatgttttt gtgacctagc ctacagtagc    24780
tacaattcgg caaaaagaaa cttttaactt aaaaacagca tactctgatt aaggttggtt    24840
acataatttta ttttctgaac tgggatcctt tttagaatga ataggggatgc tattaataat    24900
catgccctga cagcattgtg gtcaggacct gtaatcaact taactttaat taaatagcat    24960
caccatttta aaagacttga gcatgagcca cgtgcggtgg cacacacttg tagtcccagc    25020
tgcttgggag gctaaggtgg aaggattcct tgaacacagg agattgaggc cagccttggc    25080
atcatagtga gacctggtca cgtgtttggg ttaaaaatca ctaacttcaa cttctatttt    25140
ctcaatgggt aatgtcccct agatagggtc cctagttatt attaagtagg taaaaataaa    25200
ggcttgttaa tggatttagg taattatgga ggaatgagtt tggcttctgt gctttatttt    25260
accatattga ttatttgtaa tatggccatt aatacatttca ctgtttagtc tttttttgttt    25320
tacttttttat gttttactca aaatgagtgg gtgggtggaa ttctaatttt tattgttaag    25380
ggaagacatt ttaccttgtc tttaattttt tattttttt taccatttcc ctgccagtta    25440
gagatactat actatactgt cttgaatcct ctgtaggaaa acatggcata gaaataatta    25500
aataataatt agatgttaaa taataatgct gtatgactaa agaacctcct taccccacct    25560
tttctgttgt tgttctgtag taccacaata atcacttgtt aatttttattt atttatttat    25620
ttgagaggga gtctcactct cgcccaggct ggagtgcggt ggcgcaatct cagctcactg    25680
caacctccgc ctcccaggtt caagcgattc tccggcctca tactcccaag tagctgggat    25740
tacaggcgcc tgctaccaca tccggctgat tttttgtata tttagtagat acgggcttc    25800
atcatgttgg ccaagctggt ctcaaactcc tgacctcagg tgatccacct gcctcagcct    25860
tccaaagtgt tgggattatg ggtgtgagcc actgcaccca gcctcacctg ttaatttttat    25920
gagcaaaaca gattagttgg gcaagtcctt cacatgcata tctcgttgtt gttgttttttt    25980
ttaagacgga gtcttggtct tactccccaa gcaatggtgc gatttcgtct cagtgcaacc    26040
tctgcttcct gggttcaagc tattctcctg cctcaacttc ccaagtagct gggattacag    26100
gcgcccaacc acaccagct aatttttgta ttttttagtag agacagagtt tcaccatgtt    26160
ggccaggctg gtctcgaact cctgacctca ggtgatccac ccgcctcagc ctcgcaaagt    26220
gttgggatta caggtgtgag ccactgcgcc tggccacatg tgtatatctt aaaggaaata    26280
atgctaggta atttagtcag gtgcttgatg agcatttgtc atcatcatac ggagtcaatt    26340
tgtctttttc tataaaagtt cttttttgtaa atgattagtt gcttaactgc tttaatttct    26400
tctaggtacc ttacctgtca tcaggatctt tttaccacaa ataagaaacc ttaaagcata    26460
aaacttggtt cagtcttcac atttcttaaa tggaaggaga aagggaataa tagggagata    26520
tctttatctc aaagccaact gttgttgact tttccagtgg caggggtatg atgctaggac    26580
ttcagatttc ctgattcccc atcctagtgc cccttttgcc aaaactaggca ggctttcaca    26640
gcttttggag cctaatttaa gttttttcttg ttaagaacag aaaactccat tcatagattt    26700
taatttgtca ttatttgctt atttactgaa aaaaaaaaaa aactattgaa aacaggtgtg    26760
ggttaagttt cccaaaatta ggtttatatt tcaaaacaat tctaagttcc caaggataac    26820
accaaactaa gaggataaat ttttatttta tttttttta ttttatttt ttgagatgga    26880
gcctttccct gtcgcccagg ctgaagtgca gtggcatgat ctcggctcac ttgcagcatc    26940
cacctcccag gttgaagcaa ttcttctgcc tcagcctccc aagtagagaa aatcttgata    27000
agatttcaga ggcttataag gaacagcagt aaagatggt gtttttaaag ctagaataca    27060
gggatttttt taagacctt aagagaacaa acagggtttc aaaatagaaa ggagtaattc    27120
ttgtttggag aagtgaagac attttgggaaa agtttgactt gaggaaggcc attatgataa    27180
ggactgatgt ggtagactct gaaccctgaa gaatctggta cctttaagcct aaaggagggt    27240
agagtgagga cttgtgggaa aggtacagtt atatgggaag acaactgagt gtgaaaatct    27300
attcatgcaa ggctggctct taggctctaa tagaatcttg caagctaact ctagagcaag    27360
tcttttatct ttgggatgaa agaaatttaa ggctaaactt tatacattat aattataaaa    27420
ctataataaa tattagataa tgttctactg aagtataaac tcaactatct ggcatcaaac    27480
agtaactaag ccatgatcac accactgcca ctgtgttctg gcctaggtga cagagcaaga    27540
ccctgtctca aaacaaaatc cagtaactaa agagaggaat aaggggagcg caaggtaagg    27600
cagtacatgt ctgaaggggc agggaaagag ttctttctct acttccaact gggcacaaata    27660
aatcacattt gcccttttaga ttggaagagt gaaataagct tttcttgaac atcttttaag    27720
attggagtgt caaatattac caacttattt accaagaatt tggttttttct taaagtctaa    27780
agtgtttact attagctttc cacagggata taggagtttg cagaggttgt agttttttaa    27840
gggaacatca atgaatttc ttgatgcata tgcctgtttc taccattta cattgttaat    27900
ttgcctctaa aatgagtaac tcttacaatg gggtttaaaa cctgaagact attgattgct    27960
```

-continued

```
accttgatca ggtttggttt caagtgtgca ccctgtaaga gacaatgtgt ggtttattg   28020
ctattgtcac accagttttt tagatatttga aacctgtttc agatttgctt gaattgtgtc   28080
ttgtaagagg aaaatgttaa acttacttcc tctttgagaa cagttattta tagaagacag   28140
gaaaatatga gaatttttaa tagtatatga gagttctctg ttacccaaga aaagaggagtt   28200
tttttcaggc atttttaaag aatcataaat cttaaattct ttcactcagt tgctttgagt   28260
ctgtgacctg ttttacaatg gtgatagact gctttctgaa actatgaaat tggtcttgtt   28320
ggcagcatcc tacaaacata aaaagagctt cctgtgtgtc catgctccag tattttttgtc   28380
tatggtagtt atttcacaaa gccaagcgag ttacaaacga aataaaatag tgcttaagta   28440
aagaaactga ataggagaac atatactctc tcttctcaat ttttttattta gaaaattata   28500
taccttcaga aattgggggaa ataaggcatc agacactcag cttcatcatt ttttaaaagt   28560
ttgctgtatg cgctttatat gtatgttttt tgtgtatgaa ccatttcaaa gtaagttgca   28620
gacaagacag tttgctctaa ataaactcaa tgtctgggca cagtggttga cgcctgtaat   28680
cccaggactt tgggaggcca aggcaggtgg atcacttgag tccaggagtt caggaccagc   28740
ctggacaaca tggcaaaacc ccatctctac aaaaaaagag aaaaatgtca gctaggcatg   28800
gtgggcctgt agtcccagct actctagagg ctgaggtggg aggatcacct gagccctggg   28860
aggttgaggc tgcagtgagc catgattgca ccgctgcact ccagcctggg tggcagagtg   28920
agaccctgtc tgaaaataaa taccctcaat attatcacaa acatacagac tatatttaaa   28980
tttccatagt tgtcaggaat atgtctttttg tcgtctgtgt gtttttgttta attcagagta   29040
cattcaagat tcatgcgttg catttcattg ctatatttct ttaatctctt aattcagcaa   29100
agtcacctta tttggaaaaa gacctgattg tgaaagctga gtttttatgtg ttttcattga   29160
tcttgttctt tattccttgt caggttactt ttaataacat tgtgttgtaa ttggaataaa   29220
ttattaatac ttttaacatc ttattatgtg tgtgtgtaca tacactttat aggttactaa   29280
tcaggaaaaa gtcttggcta ggtcttaata cagtctttaa atcattgcct ttaagtgggc   29340
ttaaagtttt tcaaaaatgt tcttttttgta attctggaat cgaattaaga ttatgcctaa   29400
atctttacct tccttagcta aagcagtgtg gatttggggt tgattctgtt tttttttactaa   29460
taatgacgct ctagaactaa aagttaacga ttaattataa ggcaaaaaga aaagagcgtt   29520
cttttttttt ttctaattgc agagtagttt cctgacacta ctaaatgaat attttaaata   29580
aaacaggagt aattctgacc ctctgtgctt ttgtcttata acctgtactt acagtgggatg   29640
taatttttata ttaaagttta ggggtttttt tttttggtca acaagggcaa acacaagatt   29700
attcaaaggc ttgggagaaa tattacaaga agcaaggtat tgtttttatt agaaatgaga   29760
tgttgggctt ataattgtgg ttacagcaac aaagttcttt tttttcaaag gtcaggcagt   29820
tcctgctccg actggggctc ctccaggtgg tcagccagat tatagtgcag cctgggctga   29880
gtattataga caacaagcag cctattatgc ccagacaagt ccccagggaa tgccacagca   29940
tcctccagca cctcaggtat aatgtaattg ctaatttgtt gatttctact ccagtctgtt   30000
ttctgcatgt ttactgtttg tctgtttggg agtgtttgcc tttttaaattt ttatctggca   30060
aagtataata actatttaaa tgaagtacta cggtgtattg tttgggtttt tttgtttttt   30120
ataatgcttt ccagcatctg agtggtgaat atttctgcaa tgcctttgat tttaaaaata   30180
aattttcttc ccccagggat ttgcaaatca tgcaagaagc caccaccatt tatattaacc   30240
acttttttctt tcttaaagga ttcactcctg aattagctcc atttcaagga ttttctttaa   30300
cttttttgtgt atttcttatg tatctcttct gcacagggcc aataataaga agtggacaat   30360
acagtatttg cttcattgtg tggggggaaaa aaacctttgt taaatatatg gatgcagacg   30420
acttgatgaa gatcttaatt ttgttttttgg tttaaaatag tgtttccttt tttttttttt   30480
tttttgaaaa tgtacaaaat atctatcact actgatagga ggttaatatt tctgtgtaga   30540
aatgaaaatt ggtttgtttt tagtatttag tgtagatgta cacattccag caaatgtatt   30600
tgcaattatg tggttgatgc tttgtgtatat aaatgtactt tttcaatgta tactttcact   30660
tttaaaatgc ctgtttttgtg ctttacaata aatgatatga aacctcctgt gtcggtaagt   30720
tggatatgtg ggtatttaaa ggattcataa tttcttagca atgataaaatt aagatacata   30780
tacacaaata tataagcttt ccccatgaaa tattgagttt ttaaacactg gcatgttttt   30840
ccccccttgc agtatagtgg tagattggag gatctttttcc atttattgta ttggctcttt   30900
cagcacaagt aatcctgata tcttcatttt ttttccttct gtttgattaa aaactgcatg   30960
tgtgtacaat gatctttttgg catacttcca ttgcattaac agtgaaattt cctttttatac   31020
atgaccactg tttcagacct gtactgctgc tataacagtt aacctttctg ttcttaattt   31080
gataatactt gatttccaag actgtttcgg cataactaat tttaaacagt tttcagatag   31140
tgaatatgag tagtctaata agaacagttt ttttccatgt gaagcaactc tttcaatgta   31200
tataatgtta gtgtgtttct ttctaaattt aggatagaaa agtgaatagt gtgcaaaaag   31260
tatagctaca ttgcatctgc cattgaaaca taaatggggt atggaaacgt tcaagcttttt   31320
tttttttctt tatgcagtat agataagctt tgttttgtaa atgcacaagt ccaatcattg   31380
aatcaactta attttttttat gtacttgaag tcatttttatt actctttaac actcatgctg   31440
aagttctgat attttgttga aatccattgt tttactcttt gcatatttgt tggctctttg   31500
catattaata tattagacta catgcaaata cagtctgtct tgccattgtc tgttgaagtg   31560
caggtttgat ccagccagta tagaactagc tctgtagggg tgaggaggac tgtgctgtgt   31620
atcatccttg attgtgttcc ttcaaggagc attgcactgt aagtacatca gaatgacaaa   31680
ttgatgaact gcaacagtat cttttttgtca atgttccaca taatgcaaat gccatacgtt   31740
gtgtgaatat tatgttggaa tacagtgctg atatcttgga aaaccataac tgcctcttaa   31800
tttaacatag aataataacat agttctgtat ttttttttaaa gtgagcttaa tgggtaagta   31860
tttttttatat gctttagcta tagctaaaga aaactgatac ttaacaaagt tgaatagtat   31920
tattcactgg tgctcctaaa atattgtttt tcagtgtaaa atatgcatat cttctatatt   31980
taatatgaaa gtcttgaaat gtatcagaca gaaggggatt tcagtttgca aataatgagc   32040
aatgtagcaa ttttaacaca tttcataaat atatattttg tcattggtgg agagcaccat   32100
ttgttgtttt gaatatactt taaaggaaga ggtacaagga cataaatgtt gagattacct   32160
acaggatgga aatagcagta cagttcattg tagatatttt gaaatgtttt tgattgtttt   32220
atataaccta gagtgacttc ccttacccctt atttagatct gcatatatag ttctagtatg   32280
aagtttaata gttaaggagt tagctattttg ttatctttaa gagtagggta ttgacgtgaa   32340
caattgcagt attttgcatg atactgtttt atagatgacc ttttaggaaa gtggtgcatt   32400
tattaattga actgaagaag tagttcagtt gaattcagta tcataattca caaattggag   32460
gctgttgatt ttgattcatt taaggtttaa aatctttatt aattgcaaac agtgcaatta   32520
```

-continued

```
tttatacttc acagtgcctt cccagacctt ccaccttagg ttctgctgca aaaagcacca   32580
ggtaagcaca acctaaggac atatataaat aaatatttca atacattaat gttgtccctg   32640
tgaggttttt gtggttgtgt attcaaaggc aatctgctac tgcttcccca aaatgtattt   32700
tgttatttta tgctaccatc ttagtggaaa gtctgtaagt tgttaaagca actgtttaca   32760
tttctgggta atgtttttta ttttactttt ttttttttat taagacaaga aaatgatgag   32820
tagattgctg cagtaattga actacatcca aatctttttg tatttttcc ccaaatatag   32880
aagtgttaat attaagaaag gacaattaca cagtttcaa gatttaggaa atcacttgtt   32940
tagaaacttc aacagccttc acaatctgtt ttatatgatg gacagaaaat ttctttgccc   33000
tccaaaatta taatttcttt attttttct tattcttaaa ctataataat tcagtaagga   33060
tattatgggt tagaattttta ttatgatttt tttcttagac aaaagttata tgctgaagaa   33120
ggaaaaagtt ataaggcagt atgttttgat aaaaggcatg tgcatcagtg aaatgttaac   33180
tgtatagcaa ataacctttc ataatctgta gcaatcagta tttttctgat ttaataaatt   33240
tttaataact gacgctgcat ttaatttttt tgccagttta aaatgtttgt gtgtttttat   33300
agatgatttt aactggtaca tattttgagt taagttgaat gtatgaaagc agcatcttat   33360
cagtttttgtt tattcgattt ctaaaatgtg ctgatccttt taaaactcct gcttatctct   33420
gcaacaaaga aaaatattca aaaatactgc cttcattttc acacacagtg ctgaagatgc   33480
tgcaagcacc aaatcatagc tcataaaatc aggtcctgag atagttaccc ataaagagga   33540
atcctttgag tgtatgccat tggtgagccg atgagcatgg accatagaag ggctcaatgt   33600
agaaggtaaa attggcaaat cataattgag aaatatgaaa tgtattccca tacataatat   33660
ggtatagggt gtaatgtacc tgcttttgat cacttttcat tttaaagtgc tattcacttg   33720
atcttaaatg ttccatgaac tgttaaattt cttaagttac atagttacta caccacattt   33780
atgtgtatgt tatgttttaa tagtcaatga taggtatgta caattgataa tataaagggg   33840
ctcattgaaa cttgagagcc tgttgagttt tggttagttg taattgcat ttttataaaa   33900
aaaaatacag atagattgat gataatagat attggggcat tgtttctgtc tcatgagaat   33960
tcttttattc attaccataa gccttcactg atactataag cattattta aatgacgctg   34020
atcttaagtc tgaaataaat ggaaagcaga aaaggtgagc cagttgattt gaatgcattg   34080
gatattagtg ttagaaacaa tgtatagttt agattgaaag tgaactgact tatttagcac   34140
ttaaacaaaa atttgacaat gttttttagtt ttttttaaga cagcttagtg tggtgatact   34200
tagaattcta tggtttgatg tttcttttag aaatgagaag tatagtttta tttttttaata   34260
taaaaaatgt ttttaatact aaaactagta atttgatact agttgtttat aaacattgta   34320
aaatatatct tttaaacaaa ttatcttggt agttaattca taagggtggg tttgggtagg   34380
aatagcagag tactttcaga gggaagggg agtcattcag aagtgatagc attttatttg   34440
tttgaatact ctgccagtaa aatcagctgt acttagaaag ttatctgttg tgtagaataa   34500
tgatgtagag tttactaatc agtgaggatg tcttgttttt attttctgca aactctgcct   34560
cactttaaaa tgcattataa caatacctaa ttaaagataa ttttggctct gaaagttacc   34620
ttatttttg ttgagttagt gacttcattt ttcttgccac aatataagct tttgagggat   34680
tttttaaat tggtgctttt aataagcaaa taaatcccag ggttttatttt tcttcagtga   34740
tacccctata gaaactctta aatgtatttg cgcatatata tatatatatt ttcttatgca   34800
tgctcgatgc attttcgtcc tgagaaaaat gttctctaca gaaactaccc gtgtgtaaaa   34860
agaagattgg cttaaaatgg ctactgtgat gggaacagtg tcttagggag atgcagcttg   34920
gacttgaggt aaaattgaata ctttacaact gtggtttaga gtttgcttta atgacattgt   34980
atgtaaaagg tcacatgatt gctgtaattt tgtattcatt atggtttcct caataaatgt   35040
acattgatga ctatta                                                    35056
```

SEQ ID NO: 2	moltype = DNA length = 696
FEATURE	Location/Qualifiers
source	1..696
	mol_type = genomic DNA
	organism = Homo sapiens

```
SEQUENCE: 2
atggcagact attcaacagt gcctcccccc tcttctggct cagctggtgg cggtggtggc   60
ggcggtggtg gtggaggagt taacgacgct ttcaaagatg cactgcagag agcccggcag   120
attgcagcaa aaattggagg tgatgcaggg acatcactga attcaaatga ctatggttat   180
gggggacaaa aaagaccttt agaagatgga gatggctctt ggacaagtcc gagcagtaca   240
acacactggg agggaatgcc ctctcctttt aaagatcaac cagatgctaa gaaagttgct   300
cctcaaaatg actctttttgg aacacagtta ccaccgatgc atcagcagca aagatctgta   360
atgacagaag aatacaaagt tccagatgga atggttggat tcataattgg cagaggaggt   420
gaacagatct cacgcataca acaggaatct ggatgcaaaa tacagatagc tcctgacagt   480
ggtggccttc cagaaaggtc ctgtatgtta actggaacac ctgaatctgt ccagtcagca   540
aaacggttac tggaccagat tgttgaaaaa ggaagaccac ctcctggctt ccatcatggc   600
gatggaccgg gaaatgcagt tcaagaaatc atgattccag ctagcaaggc aggattagtc   660
attggaaaag ggggagaaac tattaaacag cttcag                               696
```

SEQ ID NO: 3	moltype = DNA length = 1968
FEATURE	Location/Qualifiers
source	1..1968
	mol_type = genomic DNA
	organism = Homo sapiens

```
SEQUENCE: 3
atggcagact attcaacagt gcctcccccc tcttctggct cagctggtgg cggtggtggc   60
ggcggtggtg gtggaggagt taacgacgct ttcaaagatg cactgcagag agcccggcag   120
attgcagcaa aaattggagg tgatgcaggg acatcactga attcaaatga ctatggttat   180
gggggacaaa aaagaccttt agaagatgga gatcaaccag atgctaagaa agttgctcct   240
caaaatgact ctttttggaac acagttacca ccgatgcatc agcagcaaag cagatctgta   300
atgacagaag aatacaaagt tccagatgga atggttggat tcataattgg cagaggaggt   360
gaacagatct cacgcataca acaggaatct ggatgcaaaa tacagatagc tcctgacagt   420
```

-continued

```
ggtggccttc cagaaaggtc ctgtatgtta actggaacac ctgaatctgt ccagtcagca  480
aaacggttac tggaccagat tgttgaaaaa ggaagaccag ctcctggctt ccatcatggc  540
gatggaccgg gaaatgcagt tcaagaaatc atgattccag ctagcaaggc aggattagtc  600
attggaaaag ggggagaaac tattaaacag cttcaggaac gggctggagt taaaatggtt  660
atgattcaag acgggccgca gaacactggt gctgacaaac ctcttaggat tacaggagac  720
ccatataaag ttcaacaagc caaggaaatg gtgttagagt taattcgtga tcaaggcggt  780
ttcagagaag ttcggaatga gtatgggtca agaataggag gaaatgaagg gatagatgtc  840
cccattccaa gatttgctgt tggcattgta ataggaagaa atggagagat gatcaaaaaa  900
atacaaaatg atgctggtgt tcgcattcag tttaagccag atgatgggac aacacccgaa  960
aggatagcac aaataacagg acctccagac cgatgtcaac atgctgcaga aattattaca 1020
gaccttcttc gaagtgttca ggctggtaat cctggtggac ctggacctgg tggtcgagga 1080
agaggtagag gtcaaggcaa ctggaacatg ggaccacctg gtggactaca ggaatttaat 1140
tttattgtgc caactgggaa aactggatta ataataggaa aaggaggtga aaccataaaa 1200
agcataagcc agcagtctgg tgcaagaata gaacttcaga gaaatcctcc accaaatgca 1260
gatcctaata tgaagttatt tacaattcgt ggcactccac aacagataga ctatgctcgg 1320
caactcatag aagaaaagat tggtggccca gtaaatcctt tagggccacc tgtaccccat 1380
gggcccatg  gtgtcccagg cccccatgga cctcctgggc ctccagggcc tggaactcca 1440
atgggaccat acaaccctgc accttataat cctggaccac caggcccggc tcctcatggt 1500
cctccagccc catatgctcc ccagggatgg ggaaatgcat atccacactg gcagcagcag 1560
gctcctcctg atccagctaa ggcaggaacg gatccaaatt cagcagcttg ggctgcttat 1620
tacgctcact attatcaaca gcaagcacag ccaccaccag cagccctgc  aggtgcacca 1680
actacaactc aaactaatgg acaaggagat cagcagaatc agccccagc  tggacaggtt 1740
gattatacca aggcttggga agagtactac aagaaaatgg gtcaggcagt tcctgctccg 1800
actgggggctc ctccaggtgg tcagccagat tatagtgcag cctgggctga gtattataga 1860
caacaagcag cctattatgc ccagacaagt ccccagggaa tgccacagca tcctccagca 1920
cctcagggat ttgcaaatca tgcaagaagc caccaccatt tatattaa            1968
```

```
SEQ ID NO: 4            moltype = DNA  length = 1935
FEATURE                 Location/Qualifiers
source                  1..1935
                        mol_type = genomic DNA
                        organism = Homo sapiens
SEQUENCE: 4
atggcagact attcaacagt gcctcccccc tcttctggct cagctggtgg cggtggtggc   60
ggcggtggtg gtgaggagt  taacgacgct ttcaaagatg cactgcagag agcccggcag  120
attgcagcaa aaattggagg tgatgcaggg acatcactga attcaaatga ctatggttat  180
gggggacaaa aaagaccttt agaagatgga gatcaaccag atgctaagaa agttgctcct  240
caaaatgact cttttggaac acagttacca ccgatgcatc agcagcaaag cagatctgta  300
atgacagaag aatacaaagt tccagatgga atggttggat tcataattgg cagaggaggt  360
gaacagatct cacgcataca acaggaatct ggatgcaaa  tacagatagc tcctgacagt  420
ggtggccttc cagaaaggtc ctgtatgtta actggaacac ctgaatctgt ccagtcagca  480
aaacggttac tggaccagat tgttgaaaaa ggaagaccag ctcctggctt ccatcatggc  540
gatggaccgg gaaatgcagt tcaagaaatc atgattccag ctagcaaggc aggattagtc  600
attggaaaag ggggagaaac tattaaacag cttcaggaac gggctggagt taaaatggtt  660
atgattcaag acgggccgca gaacactggt gctgacaaac ctcttaggat tacaggagac  720
ccatataaag ttcaacaagc caaggaaatg gtgttagagt taattcgtga tcaaggcggt  780
ttcagagaag ttcggaatga gtatgggtca agaataggag gaaatgaagg gatagatgtc  840
cccattccaa gatttgctgt tggcattgta ataggaagaa atggagagat gatcaaaaaa  900
atacaaaatg atgctggtgt tcgcattcag tttaagccag atgatgggac aacacccgaa  960
aggatagcac aaataacagg acctccagac cgatgtcaac atgctgcaga aattattaca 1020
gaccttcttc gaagtgttca ggctggtaat cctggtggac ctggacctgg tggtcgagga 1080
agaggtagag gtcaaggcaa ctggaacatg ggaccacctg gtggactaca ggaatttaat 1140
tttattgtgc caactgggaa aactggatta ataataggaa aaggaggtga aaccataaaa 1200
agcataagcc agcagtctgg tgcaagaata gaacttcaga gaaatcctcc accaaatgca 1260
gatcctaata tgaagttatt tacaattcgt ggcactccac aacagataga ctatgctcgg 1320
caactcatag aagaaaagat tggtggccca gtaaatcctt tagggccacc tgtaccccat 1380
gggcccatg  gtgtcccagg cccccatgga cctcctgggc ctccagggcc tggaactcca 1440
atgggaccat acaaccctgc accttataat cctggaccac caggcccggc tcctcatggt 1500
cctccagccc catatgctcc ccagggatgg ggaaatgcat atccacactg gcagcagcag 1560
gctcctcctg atccagctaa ggcaggaacg gatccaaatt cagcagcttg ggctgcttat 1620
tacgctcact attatcaaca gcaagcacag ccaccaccag cagccctgc  aggtgcacca 1680
actacaactc aaactaatgg acaaggagat cagcagaatc agccccagc  tggacaggtt 1740
gattatacca aggcttggga agagtactac aagaaaatgg gtcaggcagt tcctgctccg 1800
actgggggctc ctccaggtgg tcagccagat tatagtgcag cctgggctga gtattataga 1860
caacaagcag cctattatgc ccagacaagt ccccagggaa tgccacagca tcctccagca 1920
cctcagggcc aataa                                                 1935
```

```
SEQ ID NO: 5            moltype = DNA  length = 39750
FEATURE                 Location/Qualifiers
misc_feature            1246..1266
                        note = n is a, c, g, or t
misc_feature            2077..2131
                        note = n is a, c, g, or t
misc_feature            2666..2690
                        note = n is a, c, g, or t
misc_feature            3304..3345
```

-continued

```
                       note = n is a, c, g, or t
misc_feature           3869..3870
                       note = n is a, c, g, or t
misc_feature           6076
                       note = n is a, c, g, or t
misc_feature           6091..6093
                       note = n is a, c, g, or t
misc_feature           6159..6183
                       note = n is a, c, g, or t
misc_feature           6259..6277
                       note = n is a, c, g, or t
misc_feature           6869..6906
                       note = n is a, c, g, or t
misc_feature           7072..7139
                       note = n is a, c, g, or t
misc_feature           7714
                       note = n is a, c, g, or t
misc_feature           11374..11394
                       note = n is a, c, g, or t
misc_feature           11454..11491
                       note = n is a, c, g, or t
misc_feature           12562
                       note = n is a, c, g, or t
misc_feature           12699..12742
                       note = n is a, c, g, or t
misc_feature           12793..12794
                       note = n is a, c, g, or t
misc_feature           13648..13656
                       note = n is a, c, g, or t
misc_feature           14247..14261
                       note = n is a, c, g, or t
misc_feature           14474..20234
                       note = n is a, c, g, or t
misc_feature           21758..21789
                       note = n is a, c, g, or t
misc_feature           23223..23225
                       note = n is a, c, g, or t
misc_feature           24454..24512
                       note = n is a, c, g, or t
misc_feature           24563..24589
                       note = n is a, c, g, or t
misc_feature           25270..25338
                       note = n is a, c, g, or t
misc_feature           25723..25734
                       note = n is a, c, g, or t
misc_feature           25857
                       note = n is a, c, g, or t
misc_feature           26517..26523
                       note = n is a, c, g, or t
misc_feature           27589..27590
                       note = n is a, c, g, or t
misc_feature           27660..27688
                       note = n is a, c, g, or t
misc_feature           28733
                       note = n is a, c, g, or t
misc_feature           30266
                       note = n is a, c, g, or t
misc_feature           30317..30371
                       note = n is a, c, g, or t
misc_feature           30453..30768
                       note = n is a, c, g, or t
misc_feature           30833..30837
                       note = n is a, c, g, or t
misc_feature           30906..30912
                       note = n is a, c, g, or t
misc_feature           30997..30998
                       note = n is a, c, g, or t
misc_feature           31271..31562
                       note = n is a, c, g, or t
misc_feature           31697
                       note = n is a, c, g, or t
misc_feature           31707..31782
                       note = n is a, c, g, or t
misc_feature           32257
                       note = n is a, c, g, or t
misc_feature           33477..33478
```

-continued

```
                        note = n is a, c, g, or t
misc_feature            37353..37427
                        note = n is a, c, g, or t
source                  1..39750
                        mol_type = genomic DNA
                        organism = Macaca fascicularis
SEQUENCE: 5
gagatttcgt tccttcagtc tccacccctt tacggcacga tggtcgcgca agaatgtgat    60
acagcttcga cggccgccat tttctttctt tcttagctgt tagctgagag gaagtctctg   120
aacgggcggc agcggctagc gtagtgtaac catggcagac tattcaacag tgcctccacc   180
gtcttctggc tcagctggtg gcggcggtgg cggcggtggt ggtggaggag ttaacgacgc   240
tttcaaagat gcactacaga gagcccggca ggtaagtgtg gaccgcgcgg cggaatcccg   300
aaagctcacg gtaattggcc gctgactgag taggccgcta ccctttagcg catgaggaag   360
aggaaagagg tgtccttccg ggctgaaatg tgaggagaca cgtttcccct tgttggtaat   420
aaagattaga gaccagaact cagtttttgtg ttcttggtgt gtaatccact tagaaccccg   480
acgcgtgcta cgcaaaggcc tgaagtcttt ctcccgcttc tgcggcactc gtgtgtcgcc   540
agcgagctag cttagcctcc ccttttcctc gagatgaaga tcctcttcca ggggataaag   600
cgcaggtagt ttcacacaat ttaatggaag gttctggtaa tcagtttggg aaagaactag   660
ggtcggtctc ctggagccat agcaagggaa gggatttgtc gttaaagtag cctttacagc   720
tcatttccgt tccctctcgc aattaaaacc gctttccgta cctttcacct tctcacctct   780
acaaggaagg gacttgaaag ccgtcttttt ctgggcggga tttacgcgtc agtctgttct   840
aacagtcagt cccccttatt tctcaaaatg gcctcaggcc cattatacca gaggtttcaa   900
tttgaatctg cctctctgtt aagagtcgta aactgaccag acctctttgt attacgtagt   960
gcgtacattt gccctgaaga caccactttc ccagacgaaa gctgttaaaa tagtgcgagt  1020
attccaggaa aatagaagat ttcttaattt gaactttaca tttttagata gtcccctaat  1080
atatttaaaa tttcaaaatg tatggtgttg gtatggattt gcatgtaagc aaaagaattc  1140
tattctctat gacactatgc atgttgtact aggtgctgga cattttttact agtgttaagc  1200
taatggtagg tagaaaccag tgttgtgctg tgttttcatt gcattnnnnn nnnnnnnnnn  1260
nnnnnnncggt acagttaaag gaacattgag tcaaaatcag aattcataaa atccgttgta  1320
acatacctaa tgtgaacaca ttatcaacat gtacctgtac ttgtttataa ccagagatat  1380
tagaggtagt caacaaaagg ttatcattaa tattagttct gcggtttttt cacataattt  1440
aagaaaattct gaacatgttt agcagcaagc atttgtataa ctatgacaaa cactgttagc  1500
gcgtttttatt agattgattt gtaaaactta actattgtat ttattggaac caactatttt  1560
atttttactag aaccagcaca ttggattact tagtatcaaa cactgtaggt agatactgga  1620
gttttgtttt gtttttgtttg agacggagtc tccctctgtc gcccaggctg gattggagta  1680
gcgtggtgcg atctcgtccc actgcaacct ccgcctccgg cttcaagcaa ttctcctgcc  1740
tcagcctccc gagtggctgg aattacaggc gcccgccacc gtgcccggct aatttttttgt  1800
tttgtttttg ttttttaagt agagacaggg tttcaccatg ctggccaggc tggtctcgaa  1860
ctcctgacct cgtgatccgc ccgccttgac ctcccaaagt gctgcgatta cagcagtgag  1920
cccggccttt tttgttgttg ttgtttaaat gcatgaattg tttcctacta agaagctatg  1980
atatagttcc ttgaccaaat gcagatgaac aggattatct gataaattga ttaacgagag  2040
cagacaaaat acgggtattt aattcatctg ctcattnnnn nnnnnnnnnn nnnnnnnnnn  2100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ngttgggagg ggatacccag cttcctaagt  2160
aaaggagtga tatttgagca aggcattata gaatanaatga gagtttggcc agtagggaaa  2220
agaagcagca tgcggagagg tatgggacgt gggggaatga caagaaccca gaattgaagt  2280
cactagaaga tggtggtggg aggacgtgaa acaaggcaag gaaaggtaca agcggatcca  2340
gatactggaa gactttgtgt tgcaatgctt taatgaaaaa ttggatttcg ggccgggttc  2400
ggtggctcat gcctgtaatc ccagcacttt gggaggctga ggcagacaga tcacttgaag  2460
tcggggttta agaccagcct ggccaacatg gtgaaacccc atctctacaa aaaatacaaa  2520
aggtagccgg gcttggtggc ctgtgcctgt agttccagct acttgggagg ctgaggcaca  2580
agaattgcct gaacctggaa agtggaggta gcagtgagtg gagatcatgc tgctgcactc  2640
cagcctgggc atagagcgag accctnnnnn nnnnnnnnnn nnnnnnnnnn ggtttttatta  2700
ttaggctaag ccataatttc tttctacttt ttaaaatca gaaattgcca tcaaactgaa  2760
atttattgtc ttatatgcta ataggtcagt tgttctgtct ttactgtaat gatttctttta  2820
tgaaaatgat cttaaatctg agattcctaa ctttggcttt taactagaat tacccattac  2880
atttgatgat gtctttaaat gtcagttgta gctattagtc tgataatatg tgtggaatat  2940
acagaagggt atttagacag aagttaggtt aacatctgaa ctacttcctc cttgagtatt  3000
taagagaata ttgagttagg tttctagaat cctcaactaa ctctaagttt attttctttg  3060
tctagaatac tatgctgttt ttgttttttgg aaggaagaga tataagaaca gtttgctgct  3120
ctcaaggagc ttcaaaggct gtaccagtgg ggatgccatt ggtatattta gctggatagt  3180
tgttattcag aaaagcagga caagtaattt tgattcctgg tccgtacttg gtaatgtcag  3240
taatgttaac tctagctggt tgttgacatc tggtcattta gttgccaatt cctttttttt  3300
tttnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnntattc tgagctcaat  3360
tcaacacagg gacaccaggt tgctgctttt gttcatggtt tccagcctgt gcacttaaga  3420
aatttattttt tatgtatttc aagctgtaac tccagagatt gggattgttt gattgcgtcc  3480
ttagcagtga tactaatagc aacttctgtc tctagaacat tggaaaatta aaatgtattt  3540
atctaccgtt ttttcccttg aggttatatg aaggtagaaa tgaatcagac tagatgatta  3600
gctaagcaaa actgttaacc ctcatccctt ccctctaga caactatgaa attagtcagt  3660
atgtattcga tccttcttgc aatctcttct ctgacaatta taaagtgat tcaggctgca  3720
taatgttgtt tgaatgaaat gaaaatacag actagagctg ttttggttt tgttttttttt  3780
tttatttacc atcagtctga tcagtgaaaa ctaacattta agcatgattc tttttaaaaa  3840
tcattttgtg acagtttagc agggcttgnn tgataagcaa actatggtat ggtaatattt  3900
ctagtgtgca cgtttcttca catgtctggt atatgggaac tctaactcca tcaggacttt  3960
gcctatagta ggtactcagc atttactgaa ttaaatcaat aaacatttt gatgaattaa  4020
agtacaagtc agacctctgt gtctgtgggc tctgcatctg caaattcagc caactgtgga  4080
tcaaaaatat tagaaaaatg gaatgacggt ccaacaatac aagtaatacg aatgaaaaca  4140
```

-continued

```
atacaactat gtacattgta tcaggtatta taagtaattt agagatgctt taagtatacc  4200
aaaggatttg cataggttgt atgcagatac tgtaccattt tgtgtaagga acttgagcat  4260
ctgtggattt tggtatttgc atggttcctg gaaccaatcc ctcagggaca ctgagggact  4320
atagttggtc ataccacctg attttagaga ttttctgatc ctcagaagtt aattaagtaa  4380
actacagtag tctgttctta acctcggagg atacattcca agaacctcag tgaatatctg  4440
aaaccacata tagtattgaa tccgatatat acacggtaat attttttcct atacatgtgt  4500
atctataaag tttaaattct aaattagaca tagtattaac aataataata aattagaaaa  4560
agactgggca tagtggttca cgcctataat cttaacactt taacttttaa ctatgacgtt  4620
gtcttgaaaa agaaatcagc cagccaaggt ggctcatgtc tgtaatccta gtgtttttgag  4680
aagctcagtc aggaagtttg tttgaaccca ggaatccgag accaccctag gcaacatggt  4740
gaaaccttgt ttctataaaa aaaaaccaaa aaattagcca gacctggtgg tgcatgcctg  4800
tagtaccagc tacccaggag gctgaggtgg gaggagtgct tgaacttggg aggtcaagtt  4860
tacaatgttg acaatgttgg gtcctttacg tagttgtgta agtgagccat gatcatgcca  4920
ctgcactaca gccttgggca acagcctgac cctgtctcaa aattttaatt taattttaaa  4980
aatgaaatag aacaattaca acaatacgct gtattactga caagaagggc aaattttaa  5040
aaaaccaata tgctgtaata agttatatga ataagggat cctcccctac cccagaatat  5100
ctgattgtac tataccgtag gtaactgcaa ccgtggagag caaaaactga agatactgtg  5160
tcttaagttt cttttttcaac tcccaaattc ttggatttct cacgtcttgg cttcctcagt  5220
agaggtgaga aatgctaaaa cagtgaaaac aggaaaaata acttactcat tcaagaagtc  5280
gattatggtc cagatggaaa atttgaatta tttttgtaaa actaaactaa agtagccagg  5340
caccgtggct cacgcctgta atctcagcac tttgggaggc tgaggcgggt ggatcacttg  5400
aggtcaggaa ttcaagacca gcctggccaa caaggtgaaa ccctgtctct actaaaaata  5460
caaaaattag tcaggcatgg tggcgggcac ctgtagtccc ggctacttgg gaggctgagg  5520
tgggagaatc gcttgaacct ggggtgtgga cattgcagtg agcccagatc acgccactgc  5580
actccagcct gggcaacatt gtgagactac aaaaaaatata atagtaagta aagtaaaaag  5640
tttcccatac ttgataaatg tctaataaaa attgaatatg ttctaggact ctgaaaaagg  5700
agttgaatat agttggaggt tggttttttag gaattatttt tcttaaatta attatccttg  5760
tagtcaccta ggaattgtat attttctgtt gatcttagaa aattgatcaa atctatagtt  5820
cattttgttt tttcaatttt tttttaaaga gatggggtct tactgtatta atgttgaact  5880
cctggcctca accaggtctc ccacctcagc ttctgaagta actgggattg caggtgccac  5940
tgagccaggc tcctttattg gtatttttat taaaagcttt tctctaatgt ctttgtaaca  6000
gttctcaatt tttgaaatgg tgttactcat tctttagagt aaactgtcaa ctttcattt  6060
cttttttttt tttttnttt ttctttctt nnnttgagac ggaatctcgc tctgtcaccc  6120
aggctggagt gtagtggtgg gatctcggct cactgtgann nnnnnnnnnn nnnnnnnnnn  6180
nnntctcctg cctcagcctc cagagtagct ggaattaacg ggcacgtgac accgtgccca  6240
gctactttt tttttttnn nnnnnnnnnn nnnnnnngat gggggtttca ccatgttggt  6300
caggctggtc tcgaactcca gacgtcaggg gatccaccca cctcagcctc ccaaagtgct  6360
ggtattacag gcatgagcaa cagcgcccag cctcaacttt cattttcatt tggttagttt  6420
ttgaactatt cagtgggtaa ggttgtataa atgtcttttc tctgtataga agtttcttgg  6480
agttcaagga gtgctacttt gcaaactcat agagtattta taaaagctaa ctgcagaagg  6540
tgttcatagg ctaaaccgtt tcctattctt ggtagcacca tttttctctgg cctgaaatac  6600
tttcccctcta ctattagtgt ctgtcagtgc ccagcagtgt atttactttc ctgaggaaca  6660
attcaaaatgc taagtgcttt aagacctaag ggtgaaaag cagtgttttc aggcattatt  6720
aggaaaataa gattttaatt agacacccag aaacaaaaac aggtttgtaa ttggtaaagt  6780
gaaagatggt taaagaaggt tagattgacc aaagctagag tttttctttt tttctttttt  6840
tttttttttt tttttgagac agtgtctcnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn  6900
nnnnnggct agctgcaacc tccacttcct gggttcaaac agttctccca cctgagcctc  6960
ctgagtagct gcgaggcatg tgtcaccatg ctcagccatt ttttgtgtgt ttttagtaga  7020
gactgagttt ctccatgttg gtcaggctgg tctcgaactc ttgagctcag gnnnnnnnnn  7080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnt  7140
gtacttttt ttaatcgagc aatctagttc ttgatcctaa tagtctttgt ggtgggtgtt  7200
tgcatttta gatgaggaaa agggaaatct aggagtccta ggaaatacag ttggtatatg  7260
ggaactgata cgtaattaga cttaagcaat ccatgttgaa tttgtgactt aagcacttaa  7320
ctataaattt acccctccca gttgtctgat gataatcaaa acttgaagca gttatccata  7380
ttgggatctc tttggggaat cccagtcacc aaaagttagg ttttctttaa tattttttca  7440
tggaagtttt caaatatact caaaattgaa agaattatgt aataaattct catgagccca  7500
tcacacatct tcaataatga atgtacagca ttgcagtgtg gagcttggcc tattgctgac  7560
cactcatcaa tgtggcagaa ccactccatg gttccccatg gaaatgggag ctacttcagt  7620
tctcttttta cagaaaaatt caataaatat ctactgattg tgccctactt gtgacttgaa  7680
gccaggtttg tttgttgttg ttttttcttt tttntttttt ttttttttt actagtcttg  7740
ctctgtcacc cgagctacag tgcaatggca tgatatcagt ggatatcaat ctccagctcc  7800
tgggttcaag tgattctctt gcctcagcct cccgagtagc tgggactatg ggcgtgcacc  7860
acaatacta gctaattttt gtatatagta gagatagagc ttcatcatgt tgcccaggct  7920
gctctcgaac tcctgagctc aagcagtcta cccacctcca cctcccaaag tactaagatt  7980
acaggcatga gccaccatgc cagcaaagct gggtatttct taaatttgtt cagtcaggtg  8040
caagaaatta tttgccctac tctgaaagtt aaaaatattc taagagaatt atggtttcgc  8100
agctgggggtt ttttaagact tgagctcctg gggctcctgc atatatcctt agaacaacat  8160
tgtccagtag aagtacaatg taagccatgt tttgtttaac ccagcatatc caaaatatta  8220
cccctttgc atgtgcttaa tataaaaaat taagatgatt tctattccag gttttcaata  8280
ttcagtgagt aattttacac ttagagcaag tgtcattaga gactagtcac attttgagta  8340
ctcagtaacc acatatgggc tagtgctacc ttgctagata gcatagcctt tgggtcccac  8400
aaagtgttca cattctctgt ttacacacat cttttgaagt gtcatgagag agagccagat  8460
aggatccaga ttttctttat atctgtttgg tctgtcttac ggattcctag ttgattattt  8520
ctttgttgct tcttataggg atgtatcaga agcttaataa actcatgatt attatgttaa  8580
ctgcctgaag tattaatttt agcaacaaat tgaactataa ttttaatttt tcaagtaaat  8640
tttcttatgt gatactaagt atttcatata tacatgtatt agaagtaaat gggagtttag  8700
```

-continued

```
taatgaatag tatatagtag tttgtgatta tattttctg ttttttttt ttagattgca   8760
gcaaaaattg gaggtgatgc tgggacatca ctgaattcaa atgactatgg ttacggggga   8820
caaaaaagac ctttagaaga tggaggtaag ttatactcat acgtatttta aattgttttt   8880
cagagtgttt agttgaagtg gttctcagta tttttttgtta ttctattgag acattacttt   8940
tattataaga ttgttaaaac tgtaaggtgt atattggcct aaagggggggg aagagaccta   9000
gagaaggtaa gtaaaatttc agtcaaccag aaagtttttg cttactagaa aacagtttat   9060
tttccctctt ttaggattgt gccagattga aagtttgcac agacgtcttg ttaataatat   9120
taaaaaaaca aaaatataaa ttgcttagaa gacacttcac tggtttttact catacatgaa   9180
tggattttaa tatgctgtgt tttcgtcatt tttctatttg caattacact ttaaaggtcg   9240
aaattcgtat tgtttgctac tctagtgtgt tgcaggttat accgtttttt ttaatgttct   9300
ttgcttttag tactttttgtg tttcacgttt agctgtaaac ctatagatta aaacagtggg   9360
ttcacagtgc tttgtagttt taaaaaatcat aatcattgaa gcttctgaac ttagaagtct   9420
gcatgtattt tttgttttag gtttgatgta tgagttttga tacattttct ttttacccctt   9480
ttttttaaag gaaggattct cactgaggct atagaatgta tttgtagctt ttgaccagga   9540
gaactttgtt tcctttattt aagtgtcttt catatttata gtgaggtttt taatataaaa   9600
aaaaaaaatg agctaatgat gcttcagatg ttgtatgtaa tgtattcttt ttatttttatg   9660
tgtagatggc tcttggacaa gtccgagcag tacaacacac tgggagggaa tgccctctcc   9720
tttttaaaggc aggaattttt atttattacc tgtgttcagt atgtaaacgt gaaataaacc   9780
agtggattct taaatggaca caaatatttc ttggattagt tgtctgaatt tttgctgcac   9840
aacattctga tggttaatca tttaagtttg aaggggggag gagaatgtag tactttgagc   9900
tataggttgt ctgttccaag gtatgcattg tattcatctg tgtaatggat ttagatgaag   9960
gtagtcatgt agttgccttg agtttttttg tttttttgtgt tttttttgct aaagttttat   10020
acagtgaaat gtttgtttat aaataataga acagtttagg ttgagattgc gctgtaatgt   10080
tgtgggggttt gttttttttt ttaattttttt ggggcatagc tttgtggtca ctgtccgata   10140
tactcaactt ttaatatgtc agattttttgt agtttgatac gctttttttcc ccccagtctt   10200
cagttcctga ggtggaagca tcattagcct ttagcatgtg atattttgct agtaatggac   10260
ctaaagtact ttgtcttgta tcattctaat gtgcataaca tacattaagg tcagtgattt   10320
gttaagaatc agataactat tttaatgtct gtgcatcttt tgaatgtgaa gagaatgaag   10380
tatagtttct ttttttagatt actgagtttg tgttgaattt tggcagtttt ggttcaaatg   10440
ataaaccgta ccttcagata tttcaataaa tgtttatatt tattcttgtt tgggagggga   10500
gaagctttgt acttaattgg taaaaaatta aaagcacact aattttgcag atcaaccaga   10560
tgctaagaaa gttgctcctc aaaatgactg taagtattcc tttttaaactg ggtcaaaagc   10620
tatattaaga ttttgttcat attattgggt atctttgaag ttagttggtt tatgagtatt   10680
ttggatcagc tagcctgaac ttctttgtaa atatgtacct ttttctccta ttttacaatc   10740
tgtcccatta attgtggcca ggtatatgta aggattggtt gccgaattat tttacatatt   10800
taacaactcc aattcttgat ctacgcttgt acaacttgaa aaaggaaatt atattgttct   10860
gtgccattgc taatataatg tcttcccttt cattggctct cttacccctg tcagtgccaa   10920
tataaatatt ctgtataaaa tttgactctc ttcaaggtgt tgtctgtgca ttagggaaag   10980
gttttttgaat gtttgatgtg attttgttgt taaatacata agtaacttta aattttgaat   11040
tttcgtttta ttttatttta atttttgat agcttttgga acacagttac caccgatgca   11100
tcagcagcaa aggtatagtc acaaactttt caaaaagtac tctgcaagtt ttggttgagc   11160
tgcatgtaaa aacacaacca cattggtata tattgaatat gtgtctgtat tttttggtgt   11220
acctagttca tatcactccc cttgggaagg taccataaag tgatgatttt tcttttgagt   11280
gagaaaaatt tgtgatttgg agagagagag ggaattagcc acagtttaga agaacagggg   11340
tgaattagag taactgttaa gatgacattc tctnnnnnnn nnnnnnnnnn nnnnacagtt   11400
catgccttca gactgttcca ttgatcatcc cttcttcact tgatgtatca tctnnnnnnn   11460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn natttttga catgagtctg agcatagaac   11520
cttagtttaa gccattggga gacattagac ttccatttt attaatagat tatctttta   11580
ttgtaaacaa agtatctttc actgaaggaa aaatggtact ttctgttact ttttagcaga   11640
tctgtaatga cagaagaata caaagttcca gatggaatgg ttggatttag taagtaactt   11700
gatttttaaa gttttgaaaa catgatcaaa acatacttta gaatctttca accaaaaaaa   11760
cttttttttt ctaactagta attggcagag gaggtgaaca gatctcacgc atacaacagg   11820
aatctggatg caaaatacag atagctcctg gtaatgttac attctcttgg tattttcaga   11880
gtgactagaa aagtagcttt tttttttttct tttaggtttc tagtactaat aatgctgcct   11940
ttaatctttt gacctgaagt tctatttttg ttttaaattg tagacagtgg tggccttcca   12000
gaaaggtcct gtatgttaac tggaacacct gaatctgtcc agtaagtttg aaaatctta   12060
aaatggactt aaagtaacaa cgggagaact ctttgaattt tctctctgct ctttgttact   12120
gctttatttt acactactct ttcgttgcct ccttccctcc caagtcctct gactcctttg   12180
aagtttatgc ctcatgcctt tctcaattag ggtttatcat taaacacaga aaatggttaa   12240
aacaacttca tatctactcc agtctctact tacaaagcga agtgtagcct ggaggagaat   12300
gcgcagtaat gttgactggt gacacctaaa actcagacat taagctcaag tggactgttg   12360
tgttgtctgc atttccctag ttccattcac ttttccattc ctctccaagg ctctttaata   12420
ctatatttcc ccatctccaa atcttcagca tctaacccca cctctctcct tcttaagctt   12480
atttacggag aaaatggaaa cgaatgataa gaagcttttc ttttcccta ccgctaaacc   12540
taccagcctt cattttttt cnctgttcac atagtactca ggtaattgct ttcctttgtg   12600
ttcatgtgcc taaagccagc cctttcttcc tattgaaggt tcagcctgca gttgtacttt   12660
cttctgcgtt attagttccc cccctttact agattttttnn nnnnnnnnnn nnnnnnnnnn   12720
nnnnnnnnnn nnnnnnnnnn nntgcaggtg catgccacca cacccagcta attttttgtat   12780
tttttagta ganncagggt ttcaccatgt tggccaggat ggtgtcaatg tcttgacctc   12840
gttatccacc cgcctcggcc tccccaagtc ctgggattat aggcgtgagc cactgccccc   12900
agcctgattt tcttttgtagc actttccaaa taatacgtta ttcatttgac atgttatttt   12960
tacttatttt aatgaaacga agccactcca gcaggaaccc tgtctctttg agtgttacca   13020
ccccattacc taaaatggca cttgcacatg gttgatattt aagtatttgt tgaatgaata   13080
attgtagcat atgagtaagt aaaatggtag tttaaaaatg taaataaaata aatctttag   13140
ttcttggaag aatcagttta attctgagat aactttagca ttagagttct tggaaattgt   13200
ggactattct taaaaataaa aattgtatat ctagaaaatt tattgcctaa tctctcaatc   13260
```

```
tttgacccctt gatggcattt tctttcagtt aaaagtaaaa acattgttaa agttagcatt    13320
aaggcaccta atcctgaatt ggggtaggag gagtacttgg ttacattgtt ttgtatttct    13380
ctatttgaat aaacttgggt atgctgcaac ttactattta aatattaatt tgttaacagg    13440
tcagcaaaac ggttactgga ccagattgtt gaaaaaggaa gaccagctcc tggcttccat    13500
catggcgatg gaccgggaaa tgcagttcaa gaaatcatga ttcctgctag caaggcagga    13560
ttagtcattg gaaaggggggg agaaactatt aaacagcttc aggtattgtt atgtttgtga    13620
aatggctact tttggtctgt tttgatgnnn nnnnnngtct gctccctttt gttaatatgt    13680
attattttct atgattataa caggaacggg cgggagttaa aatggttatg attcaagacg    13740
ggccgcaaaa cactggtgct gacaaacctc ttaggattac aggagaccca tataaagttc    13800
aagtaaactt aactttatac tttataaaga aagagtttgg gttgaatggg gttgggcaaa    13860
atatgcatga ataattaaaa tgtttttgaga cgtgctttct aaattagcta acttttttcta  13920
ctttagcaag ccaaggaaat ggtgttagag ttaattcgtg atcaaggtgg tttcagagaa    13980
gttcggaatg agtatgggtc aagaatagga ggaaatgaag ggatagatgt aagtaaaaat    14040
acccattcag aaatggttgt atgctaattc gtaaatatag tagtgttttc tgttttgtgt    14100
taaatagctc taacattgtt atccttttat ttcacctta tactttagaa tacagaatta    14160
agtattttat atcttgtcac cctatttgct ataaatataa aattatatgt actattatga    14220
tttgaggcag attttcagga aatggtnnnn nnnnnnnnnn ncttttttttt actttaaacc    14280
ctgagaagct agttttctta atactcagtc ttttttacat aaggtcccca ttccaagatt    14340
tgctgttggc attgtaatag gaagaaacgg agagatgatc aaaaaaatac aaaatgatgc    14400
tggtgttcga attcagttta agccaggtga gtgcatataa taatcttgta agtgttggca    14460
gcaatgagtt ttgnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    14940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    15960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn aatttnnnnn nnnnnnnnnn    16680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16920
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    16980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17160
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17280
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17340
nnnnnnnnnn atcttgtcac nnnnnnnnnn nnnnnnnnnn aattatatgt nnnnnnnnnn    17400
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn aaaaaaatac nnnnnnnnnn    17580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    17820
```

-continued

```
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   17880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   17940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   18960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19920
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   19980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   20040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   20100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   20160
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   20220
nnnnnnnnnn nnnnacaccc atttactgtt taattaattt tgtttctttg ttttgaagat   20280
gatgggacaa cacctgagag gatagcacaa ataacaggac ctccagaccg atgtcaacat   20340
gctgcagaaa ttattacaga ccttcttcga agtgttcagg tttgatagaa agttaacatt   20400
ttcatttttt gttttttatgg aaaaatactt tccttcatga aatctgaagt ttcctctata   20460
tcagagtctg cttgatgatg ctttattaat ggagaaagtt taaattgttt taaggtaaag   20520
atcttggagc aggaagaact actaccttaa gtgctacctt atttactctt taatttaaaa   20580
aaatgttatt acttatgatt atttgggccg gttcatctct acatttatgg ttcaggtatt   20640
tgagagctgg gaaaaataga cataattaat tttatcatga aaaaaagtat ttctgtgctg   20700
attttttcct tcttgtctat tcctctctat aggctggtaa tcctggtgga cctggacctg   20760
gtggtcgagg aagaggtaga ggtcaaggca actggaacat gggaccacct ggtggactac   20820
aggaattttaa ttttattgtg ccaactggga aaactggatt aataatagga aaaggcaatg   20880
tattttaaac tcttaatgtt taacacatta ttcattttttc tggacactat ttctgttgct   20940
gttgtaaaca agtggcaatg cttttttctct ggctgtgttt tagtagaaaa gcattcttat   21000
gttaatacgt aacaagtaaa acataaatga aggatctaat gtatatttat aaaaagagca   21060
ggattttaat cttactagct tctagagaaa gtgaactaag ataattatta gcataagaaa   21120
ggtcttttga cccaaaaagt tgtttgagtg ttttttgtttg tgcattttgg tttttcccga   21180
ctcatatttt aaaaatttga atgtttataa gtgtattagt ttatatttat ctgcttttaa   21240
aagcagttta ttcaaatatt ttattataat cacattaagg ttaagtttaa acataccaag   21300
taaatgtaaa tgtattttaa gagaagcatg aaatgcttcc taaaatttag atttaaagtg   21360
tagaatatta aatgaaaaat cttaatacaa tactgtcaag tagaatactg actgaccaaa   21420
ccatgtttttt aatggcctat ttaattgtga ccattttctt ctaaatagct tctagtatac   21480
ccttgaaacc tttagagaaa ttactgtctt ttattttagg aggtgaaacc ataaaaagca   21540
taagccaaca gtctggtgca agaatagagc ttcagagaaa tcctccacca aatgcagatc   21600
ctaatatgaa gttatttaca attcgtggca ctccacaaca gatagattat gctcggcaac   21660
tcatagaaga aaagattggt gtgagtatac tttaaacttt taatttatag tgtagaccct   21720
tagattgtag ttaaattaag acgtttattc gaatacannn aatgtgaaga ttttccagca   21780
nnnnnnnnnt tatagatttc atgataccta taatagatac aatgtgaaga ttttccagca   21840
atgaaaataa cctaattaaa tgtgcagtta caggttttga gaacaacctt acgtttgagt   21900
gtggatagat aggagggtgc aggcatcata ttagtgttat tgtaggattg tggaacatac   21960
ttgaaggaca cagttaatgg gaattcatta tttattaaga ttttactata ctgaccccca   22020
gcaaggcaaa caagataaat cagatgcatc ttccgctctg cagtagaaat tcgtaaatcc   22080
tagcttttgg actggctcac aaatcatctg ggtttttaaaa tgtagactat ttgggtgctt   22140
ccaactccta cacataatat acagagacag acctgactca aaatgtctgg gatagggccc   22200
agcatcgtat tttacacaga tgtttgggtt ttttttcttcc agaaaagttt tctgttagaa   22260
caaaaagtat aaaaaagcttt gactgctttt ttgtaaatga ggcagatggt gtcttactgg   22320
agttttttaac acaaagtgtg cagggagcat cttaaattat taatcagaat ttcctaggaa   22380
```

-continued

```
aaattagttt tggtgtttgc catgctgtga atggagtgca tgcaaaacaa ggttgacgct    22440
gttttctgtg tacttgggaa gacaaaaata gatgtactaa agatgcttaa ggaacatctt    22500
ataaaatgta cacaaacatt gtaagttctg ttgtatggat agtttaattg ttcaaaaaat    22560
gaagatgttt gtggagtaat gataggaggt tgatgccata gcaaaaaaaa atgaatagca    22620
catcttggtt tttgatttta cagtattagc cttacatgca tggatcagga tttctgttgg    22680
attcatggta aaacaagata gattgtttta gagaagcaat ttggtgtggt gattaagagc    22740
acggactctt ttctctgaat catttactgg gtctgtgacc ttggagaagt tgctcaggct    22800
tttctgtgcc ttggtttcct cctataaaat gaggataatt gtatctattt catagagttg    22860
tagggattaa atgaacgttc acctgtgtgt cacttaaaag aatgcctggc acatagccct    22920
aaaaaatgtt gctacttttt cagtattatt tttactattt ggaaagaata gataatggat    22980
aataaaatgta gagatgacag caattggaga cttaaagaga caagttaata aaaataactt    23040
taaaggtcat gaagtttagt ttccctattt tgcagtgaga aatttacaac aaaatatagt    23100
gttaccattt tgtccagcat gtcgcctggt tgtgttaact actcagaagg agcattttag    23160
gacagttaag tgtaatgtct ttgttggctt aacaaaacag aaatcagtta agcattatta    23220
atnnngatgt ggcatgcatg catgataagg atataaaata tcctgattta ttgaaggact    23280
taaaaaggga atgtttgtgc tgtttagaat tatgatacat gaaaggccaa aaaggatata    23340
aattattgat ctgaatgggg tttcagtgac agaaataggt tgtgaggtgg agtttggttt    23400
aatagtgaac caactagctg ttaacattat caatgaagtg ttcagaagac aattgatgat    23460
ataggaccca agcagtttgg gaatatattg tcaagatggt tgtcttatgt taggcaagta    23520
gatgtagaga gaagagctga agataaagac ctgatttctg tgattaagat gaaaaaaagat    23580
agtaaaagaa agaattataa aaggaaaggg atctactgct gtcgcagaca agctaaagaa    23640
aatttaattg ctaattttaa tttaatttaa attgcattta attgctaccg ctattgattt    23700
tagtaaattt cacatgtctc cttattatgg cagataaaat agtttttgca aaatgaatga    23760
ggaaaggaag gaaaacctaa aatttgttgt ttgtgactat aagagggtag aaatggaaga    23820
ttatttgtcc atagagttgg tttgctgatt tttctatgtc ttccttttttt tttttgtagg    23880
gcccagtaaa tcctttaggg ccacctgtac cccatgggcc ccatggcgtc ccaggcccc    23940
atggacctcc tgggcctcca gggcctggaa ctccaatggg accatataac cctgcacctt    24000
ataatcctgg accaccaggc ccagctcctc agtaagtatt gggtttagtt ctgggctttc    24060
ccccaaagat tctagttttt ggactacatt tttatactga atttctttct cagtggtcct    24120
ccagccccat acgctcccca gggatgggga aatgcatatc cacactggca gcagcaggct    24180
cctcctgatc caggtaaaaa gatgcttatc atttgtgtgt tagctgtatt gtttttcact    24240
cgtgttacat tattaaattt tctagtgttg attctacatt tgtatgcctc accttcactc    24300
actctactct ttcaacagcg ttaggcactg cctctacccc agtgtatata gaactgacat    24360
gaatatgatc tctgctttta tgaaattcct ttcagcttgc atttggcttt cttatgagtc    24420
gatatagcag aatgataaaa actaaaagct gcannnnnnn nnnnnnnnnn nnnnnnnnnn    24480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnaatatgta gtgtgttaat tgtgacatta    24540
tctatgaaaa aatagttttt acnnnnnnnn nnnnnnnnnn nnnnnnnnng ggcactgagg    24600
ctcacgtctg taatcccacc tgttcagaag tcgaggctga ggcttgagac cagcctaggc    24660
aacatggcaa gacccatctc taaaaaaatt tgttttaaat tagccagcca gggtggtatg    24720
tgtctgtatt tctagctgct cagaaggctg gaggggggc tgaggctaga tgaggctaga    24780
agttctagat tgtattgagc tgtgggcgca ccactgcact cccctgggca acagaatggg    24840
accccatctc ttaaaaaata tatatatatg tatatataaa acagtcttac agagtgttaa    24900
gtatttgggc attaaaactcc caaattgtca aataagatac cggttctagt actcctcata    24960
atgacaactt catgtgagta aaaatcaggc ctgtatttaa taactgcatg ctaaaaccca    25020
aatacattta attattttct atattcagat ctgtatttttc gatatgtatc tattatatcc    25080
tgtactattc tggcgtctttt ccataccctg tagtttgctt tccatcttgg tcaaagagtc    25140
attctttgaa accagttaac atttatttat gatctttttt tttcacctgc tacctacctt    25200
ttaggcctct ttcaccagag tagtttcaag gaaaatatac tcaattatgg attactttct    25260
acacataatn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    25320
nnnnnnnnnn nnnnnnnntg ctgtttggaa tatgatgaca cagcatgata gttaccaaat    25380
aatcacaact tccattaatt caaggaatga cttgagttgt gctaggatat aggtatgcaa    25440
agttggggtt gtattctata gcaaaagaat gcactcccag cataccaaca ctgatagaag    25500
ggctcacaag gtacagaata taatactgta aacacgaatt tttggagaat gaaagggctt    25560
tgctttttcct cttgttggct aattgggatg gtataattaa gtaatgagat tgaagagttt    25620
gagtaggtta agagacagat tcatactggg taacttggat ctgccaggat tgtattttttg    25680
agcactactg ggtggttagc atgattgaga aaaaatgtag ggnnnnnnnn nnnaagtgg    25740
tctttgtatc acatgttgag tttcccactt ggagtagtag tgaagtcact actataaaag    25800
ctggtcagtg aatgtggttg cagcatggcc tttgggcaag aagtaaccca tttaaancca    25860
gctggttggc cccactcaga tttatcaaag ggttactggg tctctggggt ggatattgct    25920
tatattagac ttagaatgac gtaacgtttt aatgtatgaa ctaaaatatt tctttaaaaa    25980
aagagtggtc tgttacggat ttatgtagtg gtcaagaatt tagacttcag agtcaaatag    26040
acctagatca gtcctagtcc tacagtttac taatggtgag atgtcaggca agttttttgaa    26100
ctcctctaag cctgttttct tatctataaa ttgatcaatg aatgaatctt gggttgagtg    26160
aatatttagt aaaattcttag tacatactag ttatttgtga gactggtact tcagtatggt    26220
ttacacgttt gggtgtagaa ataacacttc ctaaagtctg ttttatctca aattctctgt    26280
ccaggcatag tgtaaagtga aatacctaga tttcttgatt aatatatcaga taatagccgg    26340
gcgccatggc taacacctgt aatcctagta ctttgggaga ctgaggcggg cggttcactt    26400
gaggtcagga gttggagacc agcctagcca acatggcgaa aacctgtctc tactaaaaat    26460
acaaaaatta gctgggtgtg gtggcgggtg cctgtaatgc cagctacttg ggaggcnnnn    26520
nnnggagaat tgcttgagaa tcgctccact gcattccagc ctgggcaaca gagtgagaca    26580
cttcatctca aaaaataata ataaataccg ataatgacac tattgagata tgtaaacatc    26640
caggatacaa aagcagtaca ttgggcaatt gagaaaagct tggagggtgt cctaaaaaca    26700
ccttacaatt atattacctt gtagtttttct ttttctacaa attctcactc cttttttctca    26760
aacttgacta accttttgtta agcagctgag aattgctact gttcagaatg aaagcataat    26820
agaaaattta aaagttttaa ttgtatgata ttcctagtat aaaggacaga atcaagtttt    26880
tttgtggttt ctagaagatt gagaggtctc aaactgtttg cacttcagtt gatgtgggag    26940
```

-continued

```
atgagtgagg gtcagtcaag tgtagaggaa acatagctaa aagctgagac atgggcatag   27000
tgatttctga aaagtacaag cactgtgttg tggctggagc ttgggtgtta aagagataca   27060
aatgaaggga gaggtgaggc tgaaatggaa aggataagcc aggggattca gactgttaaa   27120
gatgttgaat gataggctgt acttaagctt tgaccatcct gaagattcta gagaatgaag   27180
aatttcaagt aggatgacat caagtttata acattcaatg tgaaatagga tgaaatgtgg   27240
ctaatctttt ttttaagatt tttatatttt ctcttcattg aaaataagga caaagttcat   27300
tgttctaaaa taattgtttc tttcttatgc aggttagccc ctagaaatgt ttttcttaag   27360
tcatcttcca gatagagctg tttgtgcttg aggcgaaacc aatttagaaa aaaacaaggg   27420
cacaggatgg tttgagacag agcattggat ttggagccaa aagaccttta ttcatatccc   27480
agttatgcaa ctcagtagtt tttaacctgg gaaagtcact tggtctctga gccttgtttt   27540
ctttaataaa agctgatggt aacaacaaaa ataatagaaa atataaaann tttcatgtac   27600
tttttttttaa tgttctagtt aattttggga gctgtatatt tgccagagag ctggggggggn   27660
nnnnnnnnnn nnnnnnnnnn nnnnnnnngt gagaattaca gacttcaaca tgcaaacctt   27720
gaactttcat ttattctagc taaggcagga acggatccaa attcagcagc ttgggctgct   27780
tattatgctc actattatca acagcaagca cagccaccac cagcagcccc tgcaggtgca   27840
ccaactacaa ctcaaactaa tggacaaggt aactaaagaa cttatatgtg aaagtcaaaa   27900
cttgtgcttg gaattatata tgaagtacat cactgtataa tacctaaaat ttctaacatt   27960
atttaattat aatttaagca gacttttcct tttttaaatt gttacctgga aatagtttca   28020
aactttcata aaagttataa gaataccagg agccaccca ataaccagct gttaatgttt   28080
tatcccttgt ttactctaat accgtctata tcatacaagt gtatgggtgt gtatttattt   28140
taaatatcca ttataattag ttctgaagga aactaattat atccagtata attagttctg   28200
aaagaaatca tttaagacta aactatagac acaatgtcct gttgggtact tcttcaaaaa   28260
tatataacca tcatacagcc ctccaaatca gtaagtcaac attaatatat taagctcatt   28320
atatcattat ggatttctga tttttgcagt gggttatatt ttttgaatat ttttataatt   28380
ttgataatca gactatccta gatttggcca gtcagggagt atattcaggg tggtgcctat   28440
atcctttga aatatctgtc gttattttaa gcacttccat actgtctggc acagtgagat   28500
tgttattttg tgttttatgg gctccaaccc ttaagtcagc tatttcttca aggagctctg   28560
attcctttta gtagagtatg gtagttagaa acgaggcttg actatgcttg ttgctactga   28620
ggtgtaattg cttctaactt cttttcagcag acagaactag gaaatatatt tacatacatg   28680
catacctaca tacacatgcc aaaaaacaca taaacattga tatgtgtata tanttttttaa   28740
aactatgttc atattacgac attcatttca gcattttggg gttttcaagc cttttccttt   28800
tctgtacttg tttacaaatg gtgagaaacc tggtatcctc agtgtatctc cttattttac   28860
atacatttac ttatatgttc gacataacca gtcttccaaa caggttggtt ttcttttctg   28920
tccaccacct ctgtacccc agtaccttct atctttggca ccagtaggga tgccaccacc   28980
acatagtacc ttcctcctac ccctcactgt cacattgcag gccctgcca gctcctgcac   29040
ccaaggaaac agctaaaatt gcctttaaa aactttaatt ctgtttttct gttttgtttt   29100
gttttgtttt gcttgatcaa ggagatcagc agaatccagc cccagctgga caggttgatt   29160
ataccaaggc ttgggaagag tactacaaga aaatgggtat gttttataca tttcttgaaa   29220
atacatactt aattaaaattg aaacacaagg tattctcttc agaaagagaa taattgaata   29280
aaatcactgg actcgtaaac ataccaagac agttgcaaat tatagttttt aaatttgtgg   29340
ttatatagca aggaaatatt tttctttcta attgcatttg tcaaccagtt attaattgaa   29400
actagaaatt tcctcactgg cacatacagt attaacatta ctatactttt gacaatgaca   29460
gttatatata ttagtctgag tgacataagg ttaaatttta atgtgtcagg tgaaaatgga   29520
ttgtgtgtga taccattatt tttgctgcaa gataagcagg taagaagtaa tctgtagtga   29580
gggaaagtaa ctaagtgatg gaaccagaat ctggcttcca agagggtctg agtcccaagc   29640
ttgtctccca aatttgtctc tttagggacc atttggaacc tggtactata cttctggaca   29700
aatcactatg tttcaactgc ttttgctctt taaaaattat tacataccac agctagatgt   29760
cacaaacgaa aactaaattg gtaagcttgg ttatccttca ctagcagaaa aagaatctat   29820
aggtggtagt tttgtcataa gagatccggtc tacttgggat tctcagtgta agtttcaatg   29880
ttttctttgc caaagaatgt ttctcattct ccgcagaaag aaaaatttcc agaagggtga   29940
tgattttaat cttctaggtg taaaattata tatacttgat gataaagatg ttctgcacaa   30000
ctggtttctt tttaaagaaa gaaaaaattg tttttcctca tagagtggct tcctaggaga   30060
gtcatattcc atctatttct gagactataa cagattaata tatgtttttg tgacgtagtc   30120
taggcagatc tacagtagct gaaattccgc aaaaagaaac ttttaactta aaaacagcat   30180
actctgatta aggttggtta cataatatat tttctgaaca gggatctttt ttagaatgaa   30240
tagggatgct attaatcatg tcctgncggg cgcggtggct cacgcctgta atcccagcac   30300
tttgggaggc cgaggcnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   30360
nnnnnnnnnn ncgtctctac taaaaataca aaaaattaac cgggcgcgat ggcgggcgcc   30420
tgtagtccca gctactcagg agcctgaggc agnnnnnnnn nnnnnnnnnn nnnnnnnnnn   30480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   30540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   30600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   30660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   30720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnta aaaatcttca   30780
acttctattt tctcaatggg taatgttccc tagatatggt ccctagttat tannnnntta   30840
agtagataaa aataaaggct tgttaatgga tttagctaat tactgaggaa tgagtttgac   30900
ttctgnnnnn nnttttacct tattgattat ttgtaatatg gccgttaata catttactgt   30960
ttagtctttt atgtttttact ttttttatgtt ttactcnntg agtgggtggg ttgaattctg   31020
atttttattg ttaagggaag aaattctaat tttcattgtt aagggaagaa attgtcttta   31080
ttgtctttaa tttttttttt caccatttcc ctgccagtta gagatactgt gctatactgt   31140
cttaaatcct ctgtaggaaa acatggcata gaaataatta aatataatga ttcatgact   31200
aaagaacctc cttaccccac ctttttctatc atttttctgt agtaccacaa taatcacctg   31260
ttaatttttat nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   31320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   31380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   31440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   31500
```

-continued

```
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   31560
nntcacctgt taatttttatg agcaaaatag attagttgaa aacagatgta ggttaagttt   31620
tccaaaattg gctttatatt ttaaagacaa ttctaaattc cccaggataa caccaaacta   31680
agttattttt tttttttnttt tttttttnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   31740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnggttcaag caattcttct   31800
gcctcagcct cccaagtaga gaaaatcttg atgagatttc agagagttat aaggaatggt   31860
agcaatctaa agatgggtat ttttaaagct agaatagaag gattttttat aagagactct   31920
taagagaaca agcagggttt caaagtagaa aggagtaatt cttttttgga gaagtgaaga   31980
catttgagaa aagtttgact taaggaaggc ccttacaata aggactgatg tagtagactc   32040
tgaaccctta agaatctggt accttaagcc taaaggaggg tagagtgaga cttgtaggaa   32100
aggtacaagt atataggaag acaactgagt gtgaaaatct attcaagcaa ggctgactct   32160
taggctataa taaaatctta caggctaact gtagagcaag tcttttatct tttggatgaa   32220
agaaatgtaa ggctaaactt tatagattat aattatnaaa ctataataaa tattagataa   32280
tgttctactg aagtgtaaac tcaactgtct tgaatcaaac agtaactaag ccatgatcac   32340
accactacca ctgtgttcca gcctaggtga cagagcaaga ccctgtctca aaacaaaatc   32400
caataactaa agagaataat aaggggagca aggtaaggca gtacatgtct gaaagggcaa   32460
ggaaagagtt ctttaaagtc taaagtgttt actattagct ttccacaggg atataggagt   32520
ttgcagacgt tgtagttttt taaggaaca ttgatggaat tttcttgatg aatatgcctg   32580
tttctaccat cttacatcta aaatgataa ctctcacagt gggtttaaaa cctgaagaat   32640
attgattgct accttgatca ggtttggttt caagtgtgca gcttgtaaga gaaacgtgt   32700
ggtttttattg ctgttgttac acgagggtat ttggcagttt tttaaatatt gaaacctgtt   32760
tcagatttgc ttgaattgtg tcttgtaaga ggcaaatgtt cctgtttgat taacagttct   32820
ttatagaaga caggaaaata tgagaatttt taatagtcta tgagagttct ctgttaccca   32880
agaaaagagg gtttgtttca tgcattttta aagaatcata aatcttaaat tctttcactc   32940
agttgctttg attctgtcac ctattttaca atggtgatgg acaggtttct gaaactagga   33000
aattggcctt gttggcaaca tcatacaaac ataaaaacag cttcctatac gtccatgctc   33060
cagtactttt gtctgtggta attatttcac aaagccaagt gagttacaaa cgaaataaaa   33120
tagtgcttaa gtaaagcaac tgaataggaa aacacatctc tcttctcaat ttttttattta   33180
gaaaattata taccttcaga aattgggggaa attaggtacc actcagcttc attcatcatt   33240
ttttttaaagt ttgctgtatg tgctttatat gtacgttttt ctgtatgaac catttcaatg   33300
taagttgcag acaagacagt tggctctaaa gaaactcaat gtctggacac agtgattcac   33360
gcctgtaatc ccaggacttt tgggagacca aggcaggtgg atcactttag cccaggagtt   33420
aaagaccagc ctggacaacg ctgcaaaacc ccatctctgc aaaaaaagag aaaaatnnca   33480
gctaggcatg gtgggcttgt agtcccagct actctagagg cagaagtggg aggatcatct   33540
gagccgtggg aggtcgaggc tgcagcgagc catgaatgca ccgctgcact ccagcctggg   33600
tggcatagtg agaccctgtc tgaaaataaa taccttcaat agtatcacaa aaatacagac   33660
tatatttaaa tttccatagt tgtcaggaat atgtctttcg tcgtctgtgt gtttgttaaa   33720
ttcagagtgc attcaaagat tcgtgcattg catttcattg ttatatttct ttagtctctt   33780
aattcagcaa agtcatctca tttggagaaa gacctgatcg tgaaagctga gttatgtgtt   33840
tcattgatct tgttctttat tccttgtcac gttacttttta ataccattgt gttgtaattg   33900
gaataaaatta ttgatacttt taacatctta ctactaatca ctgtgtgtgt gtacatgtac   33960
atgtacatac actttatagg ttactaatca ggaaaaagtc ttggctaggt cttaaaataa   34020
taaaatcttt aaatcattgc aattaagtgg gtttttttgtt tttaattctg gaatccaatc   34080
aagattatgc ctaaatcttt accttcctta gctaaagcag tgtggatttg gggttgattt   34140
tgttttttta ctaataatga cgctctagaa ctaaacatta tgattaatta taaggcaaaa   34200
agaaaataaa gagcgttttt tttttatttt agagtagttt cctgacacta ctaatgaata   34260
ttttaaataa aacaggagta attctgaccc tctgtgcttt tgtcttatga ttgtaattta   34320
atattaaatt taggggtttt ttttaggtca acaagggcag acacaagatt attcaaaggc   34380
ttgggaggaa tattacaaga agcaaggtat tgttttttact ggaaatgagg tgttggactt   34440
ttaattgtgg ttacagcaac aaaattcttt tttttcaaag gtcaagcagt tcctgctccg   34500
actggggctc ctccaggtgg tcagccagat tatagtgcag cctgggctga gtattataga   34560
caacaagcag cctattatgc ccagacaagt ccccagggaa tgccacagca tcctccagca   34620
cctcaggtat aatgtaattg ctaatttgtt gatttctact ccagtctgtt ttctgcatgt   34680
ttactgtttg tctgtttggg agtgtttgcc ttttaaattt ttatctggca aagtataact   34740
tatttaaatg aagtactacg gtgtattgtt tgggtttttt tgtttgtttt ttataatgtt   34800
ttccggcatc tgagtgctga atatttctgc aatgcctttg attttaaaaa taaattttct   34860
tcccccaggg atttgcaaat catgcaagaa gctaccacca tttatattaa ccagttttc   34920
tttcttaaag gattcactcc tgaattagct ccatttcaag gattttcttt aactttttgt   34980
gtatttctta tgtatctctt ctgcacaggg ccaataataa gaagtggaca atacagtatt   35040
tgcttcattg tgtgggggaa aaaaaccttt gttaaatata tggatgcaga cgacttgatg   35100
aagatcttaa ttttgttttt ggtttaaaat agtgttttt tttgttttt tttgttgattt   35160
tttttttgcaa atgtacaaaa tatctatcac tactgtagg aggttaatat ttctgtgtag   35220
aaatgaaaat tggtttgttt ttagtattta gtgtagatgt acacattcca gcaaatgtat   35280
ttgcaattat gtggttgatg ctttgtgata taaatgtact ttttcaatgt atactttcac   35340
ttttaaaagt cctgtttttgt gcttttacaat aaatgatatg aaacctcctg tgtcggtaag   35400
ttggatatgt gggtatttaa aggattcata atttcttagc aatgataaat taagatacat   35460
atacacaaat atataagctt tccccatgaa atattgagtt tttaaacact ggcatgtttt   35520
tccccccttg cagtatagtg gtagattgga ggatcttttc catttattgt atttggctct   35580
ttcagcacaa gtaatcctgg tatcttcatt ttttttccttc tgtttgatta aaaactgcat   35640
gtgtgtacaa tgatcttctg gcatacttcc attgcattaa cagtgaaatt tccttttttat   35700
acatgaccac tgtttcagac ctgtactgct gctataacag ttaacctttc tgttcttaat   35760
ttgataatac ttgatttcca agactgtttc ggcataacta attttaaaca gttttcagat   35820
agtgaatatg agtagtctaa taagaacagt ttttttccat gtaaagcaac tctttcaatg   35880
tatataatag tgtgtttctt tctaaatttta ggatagaaaa gtgaatagtg tgcaaaaagt   35940
atagctacat tgcatctgcc attgaaacat aaatgggggta tggaaacgtt caagcttttt   36000
tttttttcttt aagcagtata gataagcttt gttttgtaaa tgcacaagtc caatcattga   36060
```

```
atcaacttaa ttttttttatg tacttgaagt cattttatta ctctttaaca ctcatgctga   36120
agttctgata ttttgttgaa atccattgtt ttactctttg catatttgtt ggctctttgc   36180
atattaatat attagactac atgcaaatac agtctgtctt gccattgtct gttgaagtgc   36240
aggtttgatc cagccagtat agaactagct ctgtaggggt gaggaggact gtgctgtgta   36300
tcatccttga ttgtgttcct tcaaggagca ttgcactgta agtacatcag aatgacaaat   36360
tgatgaactg caacagtatc tttttgtcaa tgttccacat aatgcaaatg ccatactttg   36420
tgtgaatatt atgttggaat acagtgctga tatcttgaga aaccataact gcttcttaat   36480
ttaacataga ataatacata gttctgtatt ttttttaaag tgagcttaat gggtaagtat   36540
tttttatatg ctttagctat agctaaagaa aactgatact taacaaagtt gaatagtatt   36600
attcactggt gctcctgaaa tattgttttt cagtgtaaaa tatgcatata ttctatattt   36660
aatatgaaag tcttgaaatg tatcagaagg gatttcagtt tgcaaataat gagcaatgta   36720
gcaattttaa cacatttcat aaaatatatat tttgtcattt gtggggagca ccatttgttg   36780
ttttgaatat acttaaagg aagaggtaca aggacataaa tgttgagatt acctacagga   36840
tggaaatagc agtacagttc attatagata tttttgaaatg tttttgattg ttttatataa   36900
cctagagtga cttcccttac ccttatttag atctggatat atagttctag tttgaagttt   36960
aatagttaag gagttagcta tttgttatct ttaagagtag ggtattgacg tgagcaattg   37020
cagtattttg catgatactg ttttatagat gacctttag gaaagtggtg catttattaa   37080
ttgaactgaa gaagtagttc agttgaattc agtatcataa ctcacaaatt ggaggctgtt   37140
gattttgatt catttaaggt ttaaaatctt tattaattgc aaacagtgca attatttata   37200
cttcacagtg ccttcccaga ccttccacct taggttctgc tgcaaaaagc accaggtaag   37260
cacaacctaa ggacatatat aaatataaat ttcagtacat taatgttgtc cctgtgaggt   37320
ttttgtggtt gtgtattcaa aggcaatctg ctnnnnnnnn nnnnnnnnnn nnnnnnnnnn   37380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnncca tcttagtgga   37440
aagtctgtaa gttgttaaag caactgttta catttctggg taatgttttt atttacgttt   37500
tttttttta tttaagacaa gaaaatgatg agtagattgc tgcagtaatt gatctacatc   37560
caaatctttt tgtatttttt ccccaaatat agaagtgtta agattaagaa aggacaatta   37620
cacagttttc aagatttagg aaatcacttg ttcagaaact tcaacagcct tcacaatctg   37680
ttttatatga tggacagaaa atttctttgc cctccaaaac tataaatttct ttattttttt   37740
cttaaactat aataattcag taaggatatt atggggttaca attttattat gtttttttctg   37800
agacaaaagt tatatgctag aaaaggaaaa agttaataag gcagtatgtt ttgataaaag   37860
gcatgtgcat cagtgaaatg ttaactgtac agcaaataac ctttcataat cgtgtagcaat   37920
cagtattttt ctgatttaat attttaataa ctgacgctgc atttatataa ttttttttgcc   37980
agttaaaat gtttgtgtgt ttttatagat gattttaact ggtacatatt ttgagttaag   38040
atgaatgtat gaaagcagca tcttaccagt tttgtttatt caattctaa aatgtgctga   38100
tcctttaaa actcctgctt atctctgcaa caaagaaaaa tattcaaaaa tactgccttc   38160
attttcacac acagtgctga agatgctgca agcaccaaat catagctcaa taaaatcagg   38220
tcctgagata gttacccata aagaggaatc ctttgagtgt atgccattgg tgagccgatg   38280
agcatggacc atagaagggc tcaatgtaga aggtaaaatt ggcaaatcat aattgagaaa   38340
tatgaaaagt attcccatac ataatatggt ataggggtgta atgtacctgc ttttgatcac   38400
ttttcatttt aaagtgctat tcacttgatc ttaaatgttt catgaactgt taaattttctt   38460
aagttacata gttattacac cacatttatg tgtatgttat gttttaatag tcaatgatag   38520
gtatgtaaat ataaagggac tcattgaaac ttgagagcct gtcgagtttt ggttagttgt   38580
agattgcatt tttattaaaa aaatatatagat agatgaatga taatagatat tggggcactg   38640
tttctgtctc atgagaattc tttttattcat taccataagc cttcactgat aatataagca   38700
ttattttaaa tgacgctggt cttaaatctg aaataaatgg aaagcagaaa aggtgagcca   38760
gttgatttga atgcattgga tattagtgtt agaaacaatg tatagtttag attgaaattg   38820
aactgacttt atttagcact taaacaaaaa tttgacaatg ttttttggttt ttttttaaga   38880
cagcttagtg tggtgatact tagaattcta tggtttgatg tttcttttag aatgagaag   38940
tatagttttta ttttttaata tcaaaaatgg ttttaatact aaaactagta atttaatact   39000
agttgtttat aaacattgta aaatatatct tttaaacaaa ttatcttagt agttaattca   39060
taagggtggt tttgggtagg aatagcagag taccttcaga gggaaagggg agtaattcag   39120
aagtgatagc attttatttg tttgaatact ctgccagtaa aatcagctct attagaaag   39180
ttatctgttg tgtagaataa tgatgtagag tttactaatc agtgaggatg tcttgttttt   39240
attttctgca aactctgcct cactttaaaa tgcattttaa caatacctaa ttaaaaataa   39300
ttttggttct gaaaataacc ttatttttttg ttgagttagt gacttcattt ttcttgtcac   39360
aatataagct tttgagggat tttttttaaat tggtgctttt aataagcaaa tcccaggggtt   39420
ttatttttctt cagtgatacc cctcttaaat gtatttgcac atatatatat attttttttctt   39480
atgcatgctc gatgcatttt cgtcctgaga aaaatgttct ctacagaaac tacccgtgtg   39540
taaaaagaag attggcttaa aatggctact gtgatgggaa cagtgtctta gggagatgca   39600
gcttggactt gaggtaaatt gaatacttta caaactgtgg tttagagttt gctttaatga   39660
cattgtatgt aaaaggtcac atgatagctg taattttgta ttcattatgg tttcctcaat   39720
aaataaatgt acattgatga atattataag                                     39750
```

```
SEQ ID NO: 6               moltype = DNA   length = 1968
FEATURE                    Location/Qualifiers
source                     1..1968
                           mol_type = genomic DNA
                           organism = Macaca fascicularis
SEQUENCE: 6
atggcagact attcaacagt gcctccaccg tcttctggct cagctggtgg cggcggtggc   60
ggcggtggtg gtggaggagt taacgacgct ttcaaagatg cactacagag agcccggcag   120
attgcagcaa aaattggagg tgatgctggg acatcactga attcaaatga ctatggttac   180
gggggacaaa aaagaccttt agaagatgga gatcaaccag atgctaagaa agttgctcct   240
caaaatgact cttttggaac acagttacca ccgatgcatc agcagcaaag cagatctgta   300
atgacagaag aatacaaagt tccagatgga atggttggat ttataattgg cagaggaggt   360
gaacagatct cacgcataca acaggaatct ggatgcaaaa tacagatagc tcctgacagt   420
```

```
ggtggccttc cagaaaggtc ctgtatgtta actggaacac ctgaatctgt ccagtcagca  480
aaacggttac tggaccagat tgttgaaaaa ggaagaccag ctcctggctt ccatcatggc  540
gatggaccgg gaaatgcagt tcaagaaatc atgattcctg ctagcaaggc aggattagtc  600
attggaaagg ggggagaaac tattaaacag cttcaggaac gggcgggagt taaaatggtt  660
atgattcaag acgggccgca aaacactggt gctgacaaac ctcttaggat tacaggagac  720
ccatataaag ttcaacaagc caaggaaatg gtgttagagt taattcgtga tcaaggtggt  780
ttcagagaag ttcggaatga gtatgggtca agaataggag gaaatgaagg gatagatgtc  840
cccattccaa gatttgctgt tggcattgta ataggaagaa acggagagat gatcaaaaaa  900
atacaaaatg atgctggtgt tcgaattcag tttaagccag atgatgggac aacacctgag  960
aggatagcac aaataacagg acctccagac cgatgtcaac atgctgcaga aattattaca 1020
gaccttcttc gaagtgttca ggctggtaat cctggtacctg ctggacctgg tggtcgagga 1080
agaggtagag gtcaaggcaa ctggaacatg ggaccacctg gtggactaca ggaatttaat 1140
tttattgtgc caactgggaa aactggatta ataataggaa aaggaggtga aaccataaaa 1200
agcataagcc aacagtctgg tgcaagaata gagcttcaga gaaatcctcc accaaatgca 1260
gatcctaata tgaagttatt tacaattcgt ggcactccac aacagataga ttatgctcgg 1320
caactcatag aagaaaagat tggtggccca gtaaatcctt tagggccacc tgtaccccat 1380
gggcccatg gcgtcccagg cccccatgga cctcctgggc ctccagggcc tggaactcca 1440
atgggaccat ataaccctgc accttataat cctggaccac caggcccagc tcctcatggt 1500
cctccagccc catacgctcc ccagggatgg ggaaatgcat atccacactg gcagcagcag 1560
gctcctcctg atccagctaa ggcaggaacg gatccaaatt cagcagcttg ggctgctat 1620
tatgctcact attatcaaca gcaagcacag ccaccaccag cagcccctgc aggtgcacca 1680
actacaactc aaactaatgg acaaggagat cagcagaatc agcccccagc tggacaggtt 1740
gattatacca aggcttggga gagtactaca aagaaaatgg gtcaagcagt tcctgctccg 1800
actggggctc ctccaggtgg tcagccagat tatagtcagc cctgggctga gtattataga 1860
caacaagcag cctattatgc ccagacaagt ccccagggaa tgccacagca tcctccagca 1920
cctcagggat ttgcaaatca tgcaagaagc taccaccatt tatattaa   1968
```

```
SEQ ID NO: 7          moltype = DNA   length = 6825
FEATURE               Location/Qualifiers
source                1..6825
                      mol_type = genomic DNA
                      organism = Macaca fascicularis
SEQUENCE: 7
ccttagcccc acactagcgc gttgcccgga aacatcgcga gatttcgttc cttcagtctc   60
caccccttta cggcacgatg gtcgcgcaag aatgtgatac agcttcgacg gccgccattt  120
tctttctttc ttagctgtta gctgagagga agtctctgaa cgggcggcag cggctagcgt  180
agtgtaacca tggcagacta ttcaacagtg cctccaccgt cttctggctc agctggtggc  240
ggcggtggcg gcggtggtgg tggaggagtt aacgacgctt tcaaagatgc actacagaga  300
gcccggcaga ttgcagcaaa aattggaggt gatgctggga catcactgaa ttcaaatgac  360
tatggttacg ggggacaaaa aagacctta gaagatggaa atcaaccaga tgctaagaaa  420
gttgctcctc aaaatgactc ttttggaaca cagttaccac cgatgcatca gcagcaaaga  480
tctgtaatga cagaagaata caaagttcca gatggaatgg ttggatttat aattggcaga  540
ggaggtgaac agatctcacg catacaacag gaatctggat gcaaaataca gatagctcct  600
gacagtggtg gccttccaga aggtcctgt atgttaactg gaacacctga atctgtccag  660
tcagcaaaac ggttactgga ccagattgtt gaaaaaggaa gaccagctcc tggcttccat  720
catggcgatg gaccgggaaa tgcagttcaa gaaatcatga ttcctgctag caaggcagga  780
ttagtcattg gaaagggggg agaaactatt aaacagcttc aggaacgggc gggagttaaa  840
atggttatga ttcaagacgg gccgcaaaac actggtgcta acacctgtag gattacagga  900
gacccccat ataaagttca acaagccaag gaaatggtgt tagagttaat tcgtgatcaa  960
ggtggtttca gagaagttcg gaatgagtat gggtcaagaa taggaggaaa tgaagggata 1020
gatgtcccca ttccaagatt tgctgttggc attgtaatag gaagaaacgg agagatgatc 1080
aaaaaatac aaaatgatgc tggtgttcga attcagttta agccagatga tgggacaaca 1140
cctgagagga tagcacaaat aacaggacct ccagaccgat gtcaacatgc tgcagaaatt 1200
attacagacc ttcttcgaag tgttcaggct ggtaatcctg gtgacctgg acctggtggt 1260
cgaggaagag gtagaggtca aggcaactgg aacatgggac cacctggtgg actacaggaa 1320
tttaatttta ttgtgccaac tgggaaaact ggattaataa taggaaaagg aggtgaaacc 1380
ataaaaagca taagccaaca gtctggtgca agaatagagc ttcagagaaa tcctccacca 1440
aatgcagatc ctaatatgaa gttatttaca attcgtggca ctccacaaca gatagattat 1500
gctcggcaac tcatagaaga aaagattggt ggccagtaa atcctttagg ccacctgta 1560
ccccatgggc ccatggcgt cccaggcccc catggacctc ctgggcctcc agggcctgga 1620
actccaatgg gaccatataa ccctgcacct tataatcctg gaccaccagg cccagctct 1680
catggtcctc cagccccata cgctccccag ggatggggaa atgcatatcc acactgcag 1740
cagcaggctc ctcctgatcc agctaaggca ggaacggatc caaattcagc agcttgggct 1800
gcttattatg ctcactatta tcaacagcaa gcacagccac caccagcagc ccctgcaggt 1860
gcaccaacta caactcaaac taatggacaa ggagatcagc agaatcagcc cccagctgga 1920
caggttgatt ataccaaggc ttgggaagag tactacaaga aaatgggtca agcagttcct 1980
gctccgactg gggctcctcc aggtggtcag ccagattata gtgcagcctg gctgagtat 2040
tatagacaac aagcagccta ttatgcccag acaagtcccc agggaatgcc acagcatcct 2100
ccagcacctc agggccaata ataagaagtg acaatacag tatttgcttc attgtgtggg 2160
ggaaaaaaac ctttgtttaa tatatggatg cagacgactt gatgaagatc ttaatttgt 2220
ttttggttta aaatagtgtt tttttttgt tttttttttt tgtttttgca aa 2280
tgtacaaaat atctatcact actgataggga ggttaatatt tctgtgtaga aatgaaaatt 2340
ggtttgtttt tagtatttag tgtagatgta cacattccag caaatgtatt tgcaattatg 2400
tggttgatgc tttgtgatat aaatgtactt tttcaatgta tactttcact tttaaaatgc 2460
ctgtttgtgt ctttacaata aatgatatga aacctcctgt gtcggtaagt tggatatgtg 2520
ggtatttaaa ggattcataa tttcttagca atgataaatt aagatacata tacacaaata 2580
```

-continued

```
tataagcttt ccccatgaaa tattgagttt ttaaacactg gcatgttttt ccccccttgc   2640
agtatagtgg tagattggag gatctttttcc atttattgta tttggctctt tcagcacaag   2700
taatcctggt atcttcattt tttttccttc tgtttgatta aaaactgcat gtgtgtacaa   2760
tgatcttctg gcatacttcc attgcattaa cagtgaaatt tccttttttat acatgaccac   2820
tgtttcagac ctgtactgct gctataacag ttaacctttc tgttcttaat ttgataatac   2880
ttgatttcca agactgtttc ggcataacta attttaaaca gttttcagat agtgaatatg   2940
agtagtctaa taagaacagt ttttttttcca tgtaaagcaa ctctttcaat gtatataata   3000
gtgtgtttct ttctaaattt aggatagaaa agtgaatagt gtgcaaaaag tatagctaca   3060
ttgcatctgc cattgaaaca taaatggggt atggaaacgt tcaagctttt ttttttttctt   3120
taagcagtat agataagctt tgttttgtaa atgcacaagt ccaatcattg aatcaactta   3180
atttttttat gtacttgaag tcatttttatt actctttaac actcatgctg aagttctgat   3240
attttgttga aatccattgt tttactcttt gcatatttgt tggctctttg catattaata   3300
tattagacta catgcaaata cagtctgtct tgccattgtc tgttgaagtg caggtttgat   3360
ccagccagta tagaactagc tctgtagggg tgaggaggac tgtgctgtgt atcatccttg   3420
attgtgttcc ttcaaggagc attgcactgt aagtacatca gaatgacaaa ttgatgaact   3480
gcaacagtat cttttttgtca atgttccaca taatgcaaat gccatacttt gtgtgaatat   3540
tatgttggaa tacagtgctg atatcttgga aaaccataac tgcttcttaa tttaacatag   3600
aataatacat agttctgtat tttttttaaa gtgagcttaa tgggtaagta tttttttatat   3660
gctttagcta tagctaaaga aaactgatac ttaacaaagt tgaatagtat tattcactgg   3720
tgctcctgaa atattgtttt tcagtgtaaa atatgcatat attctatatt taatatgaaa   3780
gtcttgaaat gtatcagaag gggatttcag tttgcaaata atgagcaatg tagcaatttt   3840
aacacatttc ataaatatat attttgtcat ttgtggggag caccatttgt tgttttgaat   3900
atactttaaa ggaagaggta caaggacata aatgttgaaa ttacctacag gatggaaata   3960
gcagtacagt tcattataga tattttgaaa tgttttttgat tgtttttatat aacctagagt   4020
gacttccctt accccttattt agatctggat atatagttct agtttgaagt ttaatagtta   4080
aggagttagc tatttgttat ctttaagagt agggtattga cgtgagcaat tgcagtattt   4140
tgcatgatac tgtttttatag atgacctttt aggaaagtgg tgcatttatt aattgaactg   4200
aagaagtagt tcagttgaat tcagtatcat aactcacaaa ttggaggctg ttgattttga   4260
ttcatttaag gtttaaaatc tttattaatt gcaaacagtg caattattta tacttcacag   4320
tgccttccca gaccttccac cttaggttct gctgcaaaaa gcaccaggta agcacaacct   4380
aaggacatat ataaataaat atttcagtac attaatggtg tccctgtgag gttttttgtg   4440
ttgtgtattc aaaggcaatc tgctactgct tccccaaaat gtattttgtt attttatgct   4500
accatccttag tggaaagtct gtaagttgtt aaagcaactg tttacatttc tgggtaatgt   4560
tttttatttt actttttttt ttttttatta agacaagaaa atgatgagta gattgctgca   4620
gtaattgatc tacatccaaa tctttttgta tttttttcccc aaatatagaa gtgttaagat   4680
taagaaagga caattacaca gttttcaaga tttaggaaat cacttgttca gaaacttcaa   4740
cagccttcac aatctgtttt atatgatgga cagaaaattt ctttgccctc caaaactata   4800
atttctttat tttttttctta aactataata attcagtaag gatattatgg gttacaattt   4860
tattatgttt ttttctgaga caaaagttat atgctgaaaa aggaaaaagt tataaggcag   4920
tatgttttga taaaaggcat gtgcatcagt gaaatgttaa ctgtacagca aataaccttt   4980
cataatctgt agcaatcagt attttttctga tttaataata ttttaataac tgacgctgca   5040
tttatataat tttttttgcca gtttaaaatg tttgtgtgtt tttatagatg attttaactg   5100
gtacatattt tgagttaaga tgaatgtatg aaagcagcat cttaccagtt ttgtttattc   5160
aatttctaaa atgtgctgat ccttttaaaa ctcctgctta tctctgcaac aaagaaaaat   5220
attcaaaaat actgccttca ttttcacaca cagtgctgaa gatgctgcaa gcaccaaatc   5280
atagctcata aaatcaggtc ctgagatagt tacccataaa gaggaatcct ttgagtgtat   5340
gccattggtg agccgatgag catggaccat agaagggctc aatgtagaag gtaaaattgg   5400
caaatcataa ttgagaaata tgaaatgtat tcccatacat aatatggtat agggtgtaat   5460
gtacctgctt ttgatcactt ttcattttaa agtgctattc acttgatctt aaatgttcca   5520
tgaactgtta aatttcttaa gttacatagt tattacacca catttatgtg tatgttatgt   5580
tttaatagtc aatgataggt atgtaaatat aaagggactc attgaaactt gagagcctgt   5640
cgagttttgg ttagttgtag attgcatttt tattaaaaaa atatagatag atgaatgata   5700
atagatattg gggcactgtt tctgtctcat gagaattctt ttattcatta ccataagcct   5760
tcactgataa tataagcatt atttttaaatg acgctggtct taaatctgaa ataaatggaa   5820
agcagaaaag gtgagccagt tgatttgaat gcattggata ttagtgttag aaacaatgta   5880
tagtttagat tgaaattgaa ctgactttat ttagcactta aacaaaaatt tgacaatgtt   5940
tttggttttt tttttaagac agcttagtgt ggtgatactt agaattctat ggtttgatgt   6000
ttcttttaga aatgagaagt atagttttat ttttttaatat caaaaatggt tttaatacta   6060
aaactagtaa tttaatacta gttgtttata aacattgtaa aatatatctt ttaaacaaat   6120
tatcttagta gttaattcat aagggtggtt ttgggtagga atagcagagt accttccagg   6180
ggaaagggga gtaattcaga agtgataagca tttttatttgt ttgaatactc tgccagtaaa   6240
atcagctcta cttagaaagt tatctcgttgc gtagaataat gatgtagagt ttactaatca   6300
gtgaggatgt cttgtttttta ttttctgcaa actctgcctc actttaaaat gcatttttaac   6360
aatacctaat taaaaataat tttggttctg aaaataacct tatttttttgt tgagttagtg   6420
acttcatttt tcttgtcaca atataagctt ttgagggatt tttaaaatt ggtgctttta   6480
ataagcaaat cccagggttt tattttcttc agtgataccc ctcttaaatg tatttgcaca   6540
tatatatata tttttttctta tgcatgctcg atgcattttc gtcctgagaa aaatgttctc   6600
tacagaaact acccgtgtgt aaaaagaaga ttggcttaaa atggctactg tgatgggaac   6660
agtgtcttag ggagatgcag cttggacttg aggtaaattg aatactttac aaactgtggt   6720
ttagagtttg ctttaatgac attgtatgta aaaggtcaca tgatagctgt aattttgtat   6780
tcattatggt ttcctcaata aataaatgta cattgatgaa tatta          6825
```

```
SEQ ID NO: 8          moltype = DNA  length = 1959
FEATURE               Location/Qualifiers
source                1..1959
                      mol_type = genomic DNA
```

-continued

```
                        organism = Macaca fascicularis
SEQUENCE: 8
atggcagact attcaacagt gcctccaccg tcttctggct cagctggtgg cggcggtggc   60
ggcggtggtg gtggaggagt taacgacgct ttcaaagatg cactacagag agcccggcag  120
attgcagcaa aaattggagg tgatgctggg acatcactga attcaaatga ctatggttac  180
gggggacaaa aaagaccttt agaagatgga gatggctctt ggacaagtcc gagcagtaca  240
acacactggg agggaatgcc ctctcctttt aaagatcaac cagatgctaa gaaagttgct  300
cctcaaaatg actgtaacag atctgtaatg acagaagaat acaaagttcc agatggaatg  360
gttgatttta taattggcag aggaggtgaa cagatctcac gcatacaaca ggaatctgga  420
tgcaaaatac agatagctcc tgacagtggt ggccttccag aaaggtcctg tatgttaact  480
ggaacacctg aatctgtcca gtcagcaaaa cggttactgg accagattgt tgaaaaagga  540
agaccagctc ctggcttcca tcatggcgat ggaccgggaa atgcagttca agaaatcatg  600
attcctgcta gcaaggcagg attagtcatt ggaaaggggg gagaaactat taaacagctt  660
caggaacggg cgggagttaa aatggttatg attcaagacg ggccgcaaaa cactggtgct  720
gacaaacctc ttaggattac aggagaccca tataaagttc aacaagccaa ggaaatggtg  780
ttagagttaa ttcgtgatca aggtggtttc agagaagttc ggaatgagta tgggtcaaga  840
ataggaggaa atgaagggat agatgtcccc attccaagat ttgctgttgg cattgtaata  900
ggaagaaacg gagagatgat caaaaaaata caaaatgatg ctggtgttcg aattcagttt  960
aagccagatg atgggacaac acctgagagg atagcacaaa taacaggacc tccagaccga 1020
tgtcaacatg ctgcagaaat tattacagac cttcttcgaa gtgttcaggc tggtaatcct 1080
ggtggacctg gacctggtgg tcgaggaaga ggtagaggtc aaggcaactg gaacatggga 1140
ccacctggtg gactacagga atttaatttt attgtgccaa ctgggaaaac tggattaata 1200
ataggaaaag gaggtgaaac cataaaaagc ataaagccac agtctggtgc aagaatagag 1260
cttcagagaa atcctccacc aaatgcagat cctaatatga agttatttac aattcgtggc 1320
actccacaac agatagatta tgctcggcaa ctcatagaag aaaagattgg tggcccagta 1380
aatcctttag ggcacctgt accccatggg ccccatggcg tcccaggccc ccatggacct 1440
cctggcctc cagggcctgg aactccaatg ggaccatata accctgcacc ttataatcct 1500
ggaccaccag gcccagctcc tcatggtcct ccagccccat acgctcccca gggatgggga 1560
aatgcatatc cacactggca gcagcaggct cctcctgatc cagctaaggc aggaacggat 1620
ccaaattcag cagcttgggc tgcttattat gctcactatt atcaacagca agcacagcca 1680
ccaccagcag ccccctgcagg tgcaccaact acaactcaaa ctaatggaca aggagatcaa 1740
cagaatccag ccccagctgg acaggttgat tataccaagg cttgggaaga gtactacaag 1800
aaaatgggtc aagcagttcc tgctccgact ggggctcctc caggtggtca gccagattat 1860
agtgcagcct gggctgagta ttatagacaa caagcagcct attatgccca gacaagtccc 1920
cagggaatgc cacagcatcc tccagcacct caggtataa                        1959


SEQ ID NO: 9          moltype = DNA  length = 26405
FEATURE               Location/Qualifiers
source                1..26405
                      mol_type = genomic DNA
                      organism = Mus musculus
SEQUENCE: 9
gctgaaacat cgcgggattt ccgtctgcca gtctccgccc ctttacggca cgatggtcgt   60
gcaagaatgt gatagagcct cgacggccgc catcttcttc ctttcttagc agttaaccga  120
gagcggcgcc tccctccgcg aggaaacggc agcgtgtgct gtagcgtggg tatgccgac   180
tactccacag tgcctccgcc gtcgtctggc tcggctggcg gcggcggcgg cggagtcgtt  240
aacgacgctt tcaaagatgc cctgcagaga gcgcggcagg taagcgtgga ccgaccgcgc  300
ggcggcttcc cctccccgac ccccgagagc cgccggcagg caggtccgg ccctgctcgc  360
actcgcggct gccgcccgag gagatggcgt cttctcacga ggggtcccag cggctcgagg  420
ccgagcccgt tgtcgtggtc tgaagcgccc cccctccca cgcgctctcc ccgagcgtgg  480
cctccacttc cgccccgcgc gcgcggcctc gccgccccgt gcgccctgcc tgccccggct  540
ccctcggcctc cggttccgcc cgacccaggc gcgtggaagg ttctgggcgg tcaggtcacg  600
gcgagggcaa gccagggagg gccggctgcc tggctagcgg tccccgccgc ttcccgcgcc  660
gccgagctct ccaccctcgc cgtaattagc ctaacgttcc ccgcggcgtt caccagcgag  720
gctctcccg gggcgggatc tgcgtgcgat ccaccaatcc tttccccagc tccaaatggc  780
cccgggccct ttcttcccca gccttgagtt tgaatctgcc tctctcttcc caagtcgaag  840
cctccaacct ctttgtgcta cgtagggacg tactgttgcc tcgaatggca accattttcc  900
cagatgaaaa gggttggata gataagagtc cttgcatttg acacccccccc cccaataacc  960
tgaacgtcaa tgtcttcaag tagccccgc ccccgacttc tttgtagctt aaaaatgtat 1020
aatgttgcct tgccgggtgg ggtggtgctc gactttagtc ccagcactcg agagacagag 1080
cctggcaggt ctcttgagtt tgaggccagt caaaacggcc aagagggacc tgtctcaaaa 1140
aaaaaagtat gttatggaat ttgcctgtaa ggtaaaataa acccttatgg ctctgttgta 1200
gttgctagac atgagcaggt tttaaagcc agctttgact atactggaaa cttccttcag 1260
gaccgaaaga ttaccgactc aatcaaggtt tatcaagtct accataatac agcaagcaca 1320
cttatctagt gcatagatac tgaaaaaata tatttattca tacagaggga ttccatgagc 1380
aatgtggttg gtgtacatag aagtcagagg acaacttgtg ggagttgatt ctccttccat 1440
gtaggaccct ttggtcaggt gtaagttgtt aggctcaggg acaagtgtgt ctgcctgggc 1500
atctgaggaa actttttgat gcagggcctc taccgaaagc ctaggcaggc ctgaggttta 1560
aggcttcctg ctgaattgtc tctagtgtag gtgtactcca cctctcggct ttggaagctt 1620
tttaagtgtg tgatttctag tcttgtgaga taaaaatcga acacgagtga gcagacagct 1680
ttgatgataa agtaaaccac agcggttaca taaatagtca cactgtaggc aggcgctgt  1740
acaggacgct gctggaaagg gatgcctcac tgaaagatgg gtgacattca tatagtcag  1800
tgaggagtgg gagtctggcc agcatagaca agaagctgct gcttgctcta caaagatcga 1860
gggtttctgg aaggacagct ctaagaggta gtcagctaat ataggataac ataggaggag 1920
gaaaaaggtg ggaatgtgag aaaaactatg gatagattcc attgttagca agcattggct 1980
gtgataatac aagtgtttat tatagatgat agtatctgtt caccctttgg acattctggg 2040
```

-continued

```
tatgtagaaa ttccatcagg accttgcctg tgtaagtaag ggactaaggt ctcactatgt    2100
agttctggct gtcctggaac tcactgccta cctctgcctt taggactaaa ggtgtgtgcc    2160
acccatgccc agatgagatt ttgttttcaa aatgagagtc tttctgcctc accctcctga    2220
atgctggagt gacaggtctg cctacactga tcctggcaca cttactttt ctaaaaatta     2280
ttttctttat ttacattcct gcccactcac ccccagttcc tcacaccatt ccttcttcat    2340
ctccccacct cccttccctg gggcctcaag tctcaagggt taggtggatc ttctcccact    2400
gagaccagac caggcagtcc tctgctgtta cacacacaca cacacacaca cacacacaca    2460
cacatataca cacataagcc agggaacttg gaccagcccg agtgtgctgc taggcacact    2520
ttttataagt ttggtgcaac tgtgtctttg aggatgaatg tgcgcattgt taaattttcc    2580
aaaaggaaaa tttgaaagtt gctgtggacc ctgcctagga agcactagaa gtagatttgc    2640
acttccacct ttaaaactac taaagaaagt gtgggctaaa gtgtaggaga tagatgtttt    2700
cacgtcatta ccttcttaga attatcttta actccaaaat tggtgagggt gggtcagtga    2760
agaaagtggt aaatgctggt ttggtttgat tttgtctga agtaaaaaag cacagaaata     2820
cagtagagca aaaatcaagt ggaaagcttg aattatttct gtgggttaaa ataaaaacat    2880
tacttacact taatgtgtgt ttaatagaaa cattgggttc ataattgtag catgtgagaa    2940
gttaggatta gactagctgg ggctataatc agaccatgtt cccatccacc caccacccac    3000
ccccaatgag gggagagttg aaagttttgc cttggtcttt gaatgaactg tgtaccaatt    3060
tttcttttgg atgcttgtag acctattaca tgaatggagt gtgaataagc atgctttctt    3120
atggatattc cttggtcttc aagtaaactc agaagtgttc ataaaagcta attgtagaac    3180
atagccatag tctaagccgt tttcctctcg ttcaattgta cgtaacactc ccctttaac     3240
atcagttcct gtcagtccca tacagtgcac ttaaatacct aagatatagt ttacaaatgc    3300
tatgatttga aatcaaaggg tggaaaagat tttgtttcca aatgtgtggg ttgttttaaa    3360
tggggatata acataagttt aatagatcat taactatttg ttagctccct gaagtattag    3420
tggtagtttt aattattctt aataagtgaa tctttctagt gttttaaaa gtatatcttt     3480
tattgtagga aatttagtgc tgagtagtac ataaagtatt aataccctga ttagtattcc    3540
gattttcctt ttaagattgc agcaaaaatt ggggtgatg ctggtacatc attgaattca     3600
aatgactatg gttatggggg acaaaaaaga cccttagaag atggaggtaa gttatagcca    3660
acaagaataa ttatcttcca tggtattcat ttgataagat tttcagtatt gtttatcaag    3720
atgatttatc aaaatactac tatctcattgg tctcagaaga ataggatttt gaaaaagaca    3780
gacctagaaa aggtaggtat aatttgaatc tattgggaag tattattgta atagaaaaca    3840
ggttttgtta tttttttttt tttttgtcct gttataggat tatgcttgct ttgagttgaa    3900
cagactgtta atgatataaa taaagttatg ttatttggga aacactttac tgaatttct     3960
catttattga atgagtcttt gatactccct ttcattcagt cagaatccac gtctaggtta    4020
aggttggttg gttgcaggca tcccattttt atattgtttt agtgtttttgt gtgtttcata   4080
tttatcttta aacattgatt aaaactgtgg attcatagtg ttttctcttt ttaaaaatca    4140
tcacttctgc acttagaagt gtgtactttt ttattttagt tttgatgtat gtataaattt    4200
tgatattgat atatttgccc ttttgtaaaa gattctgata gcatatagga ctgactgtat    4260
ttgtggccct ttctggtaat acttggtttc cttttgttta agtatgttta tatttgcaag    4320
tgaggttttt aaatatgcaag aaaaaatgag ctaatgatgc ttcagatgtt gtgtgtaatg    4380
tactcttttt atttttatgtg tagatggctc ttggacaaat ccgagcagta ccacacactg    4440
ggagggaatg ccctctcctt ttaaaggcag gaattttat ttattacctg tgtttagtat      4500
gtaaatgtga aataagccag tggattctta aatggacaca atggtctctt ggtttatgtg     4560
tctgaatttt tgctatacaa aagccttgaa gttgggagaa taataatagt gttttgagct    4620
ataggctgtc tattcatatg ttgtatgcat tttgtttatt tgagaaaggc aatcgtgtgg    4680
ttgctttgag gttttggtgt tctgtagtga aatgcttcat ttgagaataa tagaacagtt    4740
tagattgaaa ttgtcttgta acgtgagctt tctttttttgg tatagtttta tgatcactgt    4800
cagattcact catcttttga tatgtctggt ttttatgtag tgctctctaa ttgtcagttc    4860
tagagtggat gaagcatcat taaccattag cataaggtat tcaccagcat tgaagctaaa    4920
cagtttgtgc tgtattattc caaactaaag tgtctgtgca ttgtttaaac acaagagaat    4980
gaaagaattc aggttttgtg ggtttttttt tttttttttag attttttgagt ctaggttgaa    5040
ttgtgtcagt tttgattcaa ttggtaaacc aatatatcat gtattttaat aaatatcatt    5100
attcttgctt ggtgggggtg gggagctttg tttttaattg acaaaaagta aaagacactt    5160
aattttgcag atcagccaga tgctaagaaa gtacctcccc aaaatgactg taagtacttt    5220
aagctgaggt tttgttgctt ctttaaaaat tacgtgtgtg tgtgtgtgtg tgtgtatgtg    5280
tgtgtgtatg tatgtatgta tgtatgtatg tcaagatggc ataggaattc cctgaaagtg    5340
gactacatgt gggtgctgag aacttggttc tctacaagag catgtgtcct ctttatcatc    5400
taagtcttct ctccagactg ttgtttctac tttgggtggg ttgtgagttt ggggtttggg    5460
gaatttgctt ttgagccaaa gtctcaaata tgtaattctg actgcccttg acttcacaga    5520
gatctacctg tgtaaaagta ttttttattt tgcctgatgt aattgtcatc tctgagcctt    5580
actgcctcaa accttctagt tctttctgaa ctctggtgca aactcctctc caagctgatt    5640
caatatagct tctctcttga cctctgccgg aattgctctg cttggcctca aactatcctt    5700
ctcattttct ggcttattct gtttttacat gtgtctagct tgttccctct ctgcaaccta    5760
tatatctcta caactgttcc aataagactg tcctcccacc ccctgggcta tttatttatc    5820
tcttaagtcc tcctctctat tctcctgaga gctagacata attaatccca ttcttttttgg    5880
ggtttttttt tggggggggg gggttagaca gggtttctct gtgtagccct ggctgtcctg    5940
gaactcactc tgtagaccag gctggcctcg aactcagaaa tccacctgcc tttgcctccc    6000
aagtgctggg attaaaggtg tgcgccacca acgcccggct gtttgttttt caaaacaggg    6060
tttctctgta gctctggctg tctagaactc accctggacc aggctgacct ctgcctccca    6120
agtgctggga ttaaaaggcg tgacccacca ctgcccggct ttttttcctt ttcttttcct    6180
caaagatttg ttactttata tatgagtgca ctgtagctgc cttcaggcac actagaatcg    6240
gatcctatca cattacagat ggttgtgagc caccctgtgc ttgctgggaa ttgatctcag    6300
gacctctgga agaacagtca gtgctcttaa ccactgagcc attttcccag ccctaatttt    6360
attctttgta ctccttctct gattgatcac tttgtctgct actctaaatt aaacatcact    6420
ttcaaacatg agtgcttcct tctacaaact aacttcatct ttactgtttg ggattaaagg    6480
tcattaaagg attaaagagc aaagccacac cacaactata agcacataca cccccacccc    6540
cgtaaataac acaatcttgg ggctcaccct ataatctaag tgatcaaata gcctacaaca    6600
```

```
gacttgcttt tgttttctcc tgtgcttgga ctaaggatgt gtgtgccttg ccatactcaa   6660
cagggttggt attacaaatg tctagctgga atttccttac aaatgtatat aggtttttac   6720
tatattttta aaaaagaatg attctctttt acttgggagc cacgtatata caaacactga   6780
tggctgagct gtgactataa gttaactact ctagttgtta atctgtgtgt ccctttcctg   6840
actttcttgc tactctcagt gatgattcaa atatgaatgg tccgaatttt tttcctatgt   6900
atacatcagg gaaagagttt ggagtacgtg gtatgtttgt ctcaaataaa ctcgaactta   6960
agcttgtttt tgttcatttt ttcgtagctt ttggagcaca gttacctcca atgcatcagc   7020
agcaaaggta tagtcaaaac tttaaagct actttgcaag tcttagttgg attgtatgta   7080
cagccaagtt agtctataat gcaaatgtct gcataccatt ttgtatgtgt ggtttacttc   7140
tgtaaaccac attgggaaaa tggtatagaa acattgttaa tttgggaaaa caggttgaaa   7200
ttgacgcatt tcagaagagg taaaataaat gttaaagtgg cattcgcttt gacaggtgtt   7260
cagtttcctg ccctgctcac acatgctgcc tcatttaaga acctggctgg gctactggga   7320
gactttaaga ttgggtttat tagactgctt ttatgtgtat atgtacatta tttttctgtg   7380
aggaaaatgt tagtttctgt tctttcttag cagatctgta atgacagaag aatacaaagt   7440
cccagatggg atggtcggat ttagtaagta ccattaatac taaggtttta agacaaagca   7500
tactttcaaa aatagtatga atgataactt ctccccttc tctaccttaa ctagtaattg   7560
gcagaggagg tgaacagatc tcacgaatac agcaggaatc tggatgcaag atacagatag   7620
cacctggtaa cattccgtgt gtgtgtgtgt gtgtgtacga attacttgtt gttttgtcca   7680
gttcttacag ttttttgaga gtggctagga aactcaagta ttttaagttt ctaaatatct   7740
gatatttta ttgtattttt tacttaaact tttctttta tttaattgta gatagtggtg   7800
gcctaccaga aaggtcttgt atgctaactg gaacacctga atctgtccag taagtttgaa   7860
aatgttaagt ctacttacag ttaagactga agactataag cttcgttttc ataccatttc   7920
gctttaaaag tttacatatt atgccttct taaaataggg ttcatcatct gaacctaaga   7980
cagagaaatt gttaaagcca gagactaagc cgtaactcag gcatttaggt gcttgccttc   8040
catgtaggaa gaactgagtt tgatcctaaa actaaaaaaa aaaaaactgg cgtggtaata   8100
tcatggctca ggagttggac acaggcagat tttggggact tggtaaccaa ccagctaaca   8160
ctatttgatg ggttcaaagt cattgagaga ccttgcctca aggtacttat cacaaaaaag   8220
gtactgatgg ttgtaatttg acctatacac cttccatacg tgtctataca gactcagaaa   8280
ttattaaaac ctgccttgca caaatatgtt aataacgtca cctaaaactc atgaccaatc   8340
tagtcatgaa aattttatat tcctgttttt ctcttccact caattctaag gctctgataa   8400
aactacccaa tctttacctc taattgtctt ctccccaccg catctcctcc tccagacagg   8460
gtttcattgt atatccatag ctgactgtct ctgaactcac tcgctctgta gtccaggctg   8520
gcctaaaact taaaagatct gcctgccgaa ggcatgtacc accattgccc accacactca   8580
ctgtcttgac tgacttgtgt tgctgtgacc aagaaagtgt tatcatccta cactgaaatc   8640
ttaacaccgt tccttgcata cttttttttt tttgtctagt attgagatac tgttttacgt   8700
tcatttgcct gactttctac ttcaagaaat tcctgactaa actgattttt tacttcattc   8760
ctaaaatttt aggagtaatt tctaagagtt tatagacaga tttggaaatt acatatcata   8820
aaaatatctt tgaacctcat gttgtatttt gttaacacag tgctttaaaa ggtttgtacc   8880
attgaagttt cccctttaaa agtttatcta gtgtgtgtgc ctctcagttt ccttcaagct   8940
ctgcttctaa cagctgcatc gacagttctc tttccaaggc agaggctgag tcgtgcacta   9000
gttaaccccg ttgctgtgtg acaggatgga tggaagatgt ttaatgactg tgttcacttg   9060
gagcatacag acttcagtac ttactgccca tttgacttga cactcactcc tttggtttag   9120
gttcttgcct ccgaccatga aatgccatat gaccaattta gagtatgttt gaaacctgac   9180
taagataaaa ctgctgactt ttattatttt gggataagtt gggttatgct acgtatttt   9240
atcaagatta atttgttaac agatcagcaa aaagattatt ggaccagatt gttgaaaagg   9300
gaagaccagc ccctggcttt catcatggtg atggacctgg aaatgcagtt caggaaatca   9360
tgattccagc cagcaaagca ggactagtta ttggaaaggg gggcgagact attaaacaac   9420
ttcaggtatt attatattgt aaagaatatt attaagtcat ttgaggccac tttttattgt   9480
ttttactaac atattttcat gatattaata ggaacgggct ggtgttaaaa tggtaatgat   9540
tcaagatggg cctcaaaaca ctggtgctga taaacctctt aggattacgg gtgacccata   9600
caaagttcag gtaagcttca ctttgtctct tatatataaa ggctgagggg gtggacaaaa   9660
tacgtattaa atttttttga caaatattct gaattggata actttctatt atagcaagcc   9720
aaggaaatgg tattagagtt aattcgtgat caaggtggtt tcagagaagt gcggaatgag   9780
tatggctcaa gaataggagg caatgaaggg atagatgtaa gtaagaattc tgagtcagca   9840
gtggttgtat cctaatccat gggcactagt gttttcctgc tttgaattaa atggacttgg   9900
cttaacatca tgtccctttt cttgtttct tccttggggt tttttatct tgttattttt   9960
atgatcattt agtatatta ctttatgat taagtggga tgagtttttg agaaatgtct  10020
tttaaaatat actttactt aaggtcccaa ttccaagatt tgctgttggc attgtaaatag  10080
gaagaaatgg agaaatgatt aaaaaaatac aaaatgatgc tggtgttcga attcagttta  10140
agccaggtaa gtttaaggta tatttaatgt tctaagaatt gatagctttt aatcttgaca  10200
tttataattg gtttgtcaca acttactctg tttttttttg ttgttgttgt tgtttattgt  10260
ttagatgatg gaacaacacc tgataggata gcacagataa caggacctcc agacaggtgt  10320
cagcacgctg cagaaataat cacagacctt ctacgaagtg ttcaggtgtg ataggaaatt  10380
aacatttca gtttgctttc tttggaaaag tttttatcat acctgaatga aaaaaatgcc  10440
cctccaggtt atattagaat ctgttggcag tgttaaaatg actaacgtta aattgcttta  10500
aggtaaacat tttagggtag atttgcataa atgttctgcc ttattcccc cagtgcatgt  10560
gaactttat ttctatgggt atatgtgcct tggcaattgt gcctatggaa gccaaaggaa  10620
attttttctga agttggttct ctattatgtg ggtcctagaa attaaaccca ggtcattttg  10680
catgtatggg tagtctccca atatatttta aaagtggaag ttattactta gttgactatt  10740
tttgtttgac gtcatctctt aatttcattg tttagaaata cgagagttgg ggataataag  10800
ctaatagttt tattacagag tttattttgt tctctccttt cttaggctag caatcctagt  10860
ggaccgggac ctggtggtcg aggacgaggg agaggtcaag gaaactgaa tatgggggccc  10920
ccgggtggac tccaggagtt taatttcatt gtgccaactg gaaaactgg actgatcatt  10980
ggaaaaggca atgtatttta aactctttgt tttgatatac tggatatttg taaacaaaag  11040
atgctatta tattttttta ggaaagggg cttaacatgg cagctaaaat aagaatactc  11100
tttaatgtat gctaaaaatg ggattgtgtc tcaactgctt ctagggaaag gggaccaaaa  11160
```

-continued

```
gtctgtcagt tacaaaactt gatcctaatt caaatgggtt ttgttggtgt ttttttgtttt   11220
gttttgcatt taacttcttt gttacttaca aaacctgttg attgaaatat ttttctgttc   11280
tgacccatat cagggtgtat gtttagtaag gtgtgtataa aatagacttt agtgtgtttt   11340
gcattagttc tagatggaag aacaggctaa atagaaaatc ctaacattat tgtccagtgg   11400
aatgctacaa aactgtttaa tgggcaaagt aataggtgtt tggggttttt tttggtttt   11460
tttaaatagt atttctgcca gacatggtga cagacacctt taatcccagc acttagtagg   11520
cagatgcagg tgtttctctt gtgagtttga agccagcatg gtctacaggg tgagttatag   11580
gacaaccagg gccacacaga gacccttgtc ttaggaaaga gtattgcttg cattctgtga   11640
acttagtgtc ttgcggtttt aggaggtgaa accataaaaa gcataagcca acagtctggt   11700
gcaagaatag aactgcagag aagccctcca cctaatgcag atcccaatat gaagttattt   11760
acaattcggg gcactccaca gcaaatagac tatgctcgac aactcataga agagaagatt   11820
ggggtaagtg tactctcatt tatagcatag ttcctggtaa cttaaaatg tactgtggga   11880
aaatacatcg tttctaatat gaaattgcag gtattgtgat atagtgtata tacatatatt   11940
gacataggaa agatatccaa aatacaagaa ccatgaactc agaaatccgc ctgcctctgc   12000
ctgtcaagtg ctgggattaa aggcgtgcgc caccacaccc ggctgagaat accatgtgtt   12060
aagtttccc ttgatgtgca atttgcagaa tatactctca gcagttgtca cattagggaa   12120
gtaaagggaa ctaggcagtc ttgtttttaa aagaaaatca taaaaagagt aagaaacaca   12180
gaagactaag ctgaaagaac tctgctggag ctgaggattt cagtgaattt tacatggcag   12240
atgatgacat tttatttttt gcaccaaaat tcaaagaggg taaatttaaa tttgttattt   12300
aagcgtagaa atggaagagt agttggactt tgaattactc aaccattaca ttgtttgagt   12360
gtctagttat tgtttatgtt ttcttttcct ttttttttt ataatagggc ccagtaaatc   12420
ctttaggggcc acctgtaccc catgggcccc atggggttcc aggtcctcat gggcctcctg   12480
gacctccagg gccgggaact ccaatgggac catcacaccc tgcaccttac aatccaggac   12540
cacctggccc agctcctcag taagtaccac atttgcttgg ttctgggctt tccctaaaga   12600
cctaggaatt gggggctgga gagatgacga tggctcagca gtgcagagca cttgttcctg   12660
ttgcagaggg acctgagctc acttcccagc accacacggg ggttcgcaac cattcgtgac   12720
aggttccagg gtatctttac agcgacacta ggcatacata gctgcatata catacatgca   12780
atacatgcag aggcaaaacc tacacataaa ataaataaag ctggctttta ttaatggtat   12840
ttgtgttttg ttccccaccc ctagcggtcc tccagcccca tatgtccccc agggatgggg   12900
aaatgcgtat ccacattggc agcaacaggc tcctcctgac ccaggtaaag ggtaatctac   12960
tattaatatg ttagcttcat tgtatactgc tggtcacagt cctaaactac ctactgcatg   13020
tgcgatcttt taaggcactt ccctgcagt ataggatgac actgctgtct agtatgttgt   13080
gactttattc atagtagttc atatttata aataatgtgt ttaacactac tatgttctca   13140
aataataagt ataaattatt ttgttgctga taaaattaag gtcccttaat taccaaaaca   13200
ggagagtaca tttattttc tctcttataa tctggatttt ctatagcatg gtgttctcat   13260
tatatcttcc catactgttt ttgtttctac tccatcatga tcaggaatca acttggaaat   13320
aggacttttt tctttgtttt tcttttgttt tttcactttg cagccccctgt agtttgattc   13380
atttgaaaag tttgagagaa cgtgttctga gttatagatt tctacatagt gtattttta   13440
caaatcaaag ttgatttaga tgtaattgct ttgttggatt tgttaaatgt tttaatattt   13500
taatgtatgt taaattcata ggaacttttc tcagtcacca aataaccaca acttctattc   13560
atttaagaaa caacttgagc cgggcggcag tggtgaacgc cattagtccc agcacacagg   13620
agacagaagc aggcagatct ctgagttcga ggacagcctg atctacagag catgttcaag   13680
gatagccagg gacacacagg aaaaatccta tctcaaaaaa gagagagaga gaaaaaaaaa   13740
gaaaaagaaa taacttgaat tgtgatggca tctgaatatg aaagatgagt gtcgattagc   13800
aaaaagagtgc attcccaggg cagacaagaa gcatcagtta tgacttcctt tgtaagggtt   13860
ttaaaagtga atcttctcct tctaacaagc atttagacgt ggcatgatca aatagagttg   13920
gaacaggtta tattgaaggt tttatatggc taattcaatt tcataggatt gctcaaatga   13980
gtgatgtttg atattagcac tgtaaggtct ggtcagtgaa tatgacttcc agcattacct   14040
ggaagactga ctcggcaata tgttgcttaa atttgacttg gaagtataca gtattttatt   14100
gtctgcattt tcaagaaact gctgctgtga ggttgagatt tatatagatg tcaagaatttt   14160
aatcttcaat cttaaatgga taatatgcga attcatccta tgtaaaagaa tagtgagctc   14220
ttagtacctg ctagttgcaa gtcacaaaaa aaaagtttgg gtttggttt ggtttttgag   14280
ttacaaaggt gagcttagtg cagaggatgc tggttgtggt gacttgggaa tagggcagca   14340
gctaaaactt gagtattggg ggatgctaaa agaattaata gaaaggataa gccacagtgc   14400
tggggataag cattagaacc tgagttctat ccctcccacc tgttgtaaag aaaaataaat   14460
aaagcgcttt aataaagttc gtaaaaagga tccgtaaagt ttcctggcct gctcttgtaa   14520
ccaaatctag gagctctggg ttcagagaga aatcctatct caaaaaaaat taggaaggaa   14580
ggctggagag agggctcagt ggataagaaa agcaataact gctcttccag aggtctgagt   14640
tcaattccca gcaaccacat ggtagttcac aaccatctgc aatgggatct gatgccctct   14700
gctggtgtt ctgaagagag cttcaagtgt actcatatac aaaataataa taattctttt   14760
aaaaaatgag gaagggtgtc agagagggtag cttagtgatt attataaagg tcttcctgct   14820
ctttgagagg atgcaagttg agatccagtt actctcacac cttctggttc tctttagtaa   14880
cccacatttg tgtagataca cacaggttca cagatacaca taactagaag tgggactttta   14940
aagtctagat aaaacaaggt gagcatgtct agggagatca cccagaatca gtacacacaa   15000
acagacacgc cacgacacga tggctaagct ggggagggcg gccagaccac agcagtgtagc   15060
ccggaaacag gaaaagtgga ggaagaaaag gtaacggctt ggaaatccaa cggcctgtgct   15120
taaatgggc atggtcttaa aactgtacag gagcctttcaa atgataagct atgcttgaag   15180
tatcctaaag gtgttagaaa aggatagagc aggactttca ggattgtgac tgaaataaat   15240
ataaactgca tgtatctaac ctttgtttct ttcttgtaca aattagtcca taactgcctg   15300
ctatacagaa ctgttgcact tggcagctta acttgttggg ggaaaaaaag caaagacacg   15360
ggattgagct gagacaaacc attggattgg aagccaaagat catccatt tatataccag   15420
ttccaccgat cactagtaaa ttacttaata cctgtaacct aggaaattac ttaattcccg   15480
aggctgaagt ttctgaacta aaagccaagt gatgatgata atactagaac attttgaaag   15540
tttccattcg ctattggtaa ctttttgagcc tatttgtcag gaatcaccaa aaaggtaagg   15600
gttggtttct agagtttaca tattaacctt gaaaccgctc tttgttttta cattgtagcc   15660
aaggcaggag cagatccaaa ctcggcagct tgggctgctt attatgctca ctattaccag   15720
```

-continued

```
cagcaggcac aaccccacc tgcagctcct gccggtgcac cagctacaac ccaaacgaac   15780
ggacaaggta actacggtaa tgaaagattt tatttcctat tttaaagcta tagtttgtgt   15840
ttggattcat gtatgaagtc ataacaaaac tctctttcag tcctgagaat tgttgggcct   15900
tgtgcatgat ggtatcatta ggttagatcc tcgggcctaa ggttttgaag actcagttgt   15960
gactttgtgt aagcaaacta agtcaaataa tagttggaag aataatgcct taagtagtga   16020
gactgttaac tgtcctcccc ccacccccat ttacttcact accctgattt atgtctgtta   16080
taatagttct gaagagagct actgaaaact gtacttagag acataaaatc atatcacctt   16140
tctacagacc cttattcatt ccaaaataaa tgaccaccat ggactttaaa gtcagttact   16200
catcattgat ataccactgt tatacacagc ccttgcttag atcaccaact gaccccgtgt   16260
tttcattagg ttcattaagt ctcatcaagc tagagcactt tcttggtctt cttttatgcc   16320
tctaataaga ggataactca acagttagat agctgccaag ctatagagtt actaaccatg   16380
gttgtttaaa tggtcttgct actaccagcc tggtttttgc actatactac aactatgtac   16440
tgtaaagaat tacctagggg ctggtgagat ggctcagtgg ataagagcac ccgactgctc   16500
ttccaaaggt ccggagttca aatcccagca accacatggt ggctcacaac catctgtaat   16560
gggatacgat ccgattctag tgtgcctgaa ggcagctaca gtgtacttac atataataaa   16620
taaataaatc ttttaaaaaa aaaaagaatt acctagtatt ttctgctact ttttactgga   16680
gtaatgtcac tatcctatac attgaacttc tgcctcagcc ttagtctgta ttattaataa   16740
accttgcctt ggtaccctac tattacagca gctgttaact gttagtttct gtcattgtat   16800
attagttgct attttactac aaggaataga ttgccctttt ttaactactt tcattatatt   16860
gtcctagatt tggggtggtt ggggttatct tcaaggtggc tctggtgcct ctgacgttcg   16920
ttcttcaggt gcttccttag tacctgacac agtaagataa ctttgtcttc tgagctccaa   16980
tcggtgattc atttcctttt ctctctctct ttctctctct ctctctctct ctctctctcc   17040
ctctctctct ctctctctct ctcaggatag aatttagaga tcagtctaag   17100
tatactaatt gctatctaaa tgtcactaat tataatcctg tccaaattag aactaaaaag   17160
tatgtttggg gctggcattc agtcaatagt acttgttcct gcagaaaacc caggtttgat   17220
tgctagtacc cacatggtaa ctcacactga aactgattcc aggggaccca gtgccctctt   17280
ctgatctgtg gagcaccaga tacccagggt gcacagatta tacatgtgga caaaatgctc   17340
acacataaaa ttaaacagta tctttaaaaa catgcaggac aacacgtgtc tactaaaaat   17400
cacatgatta tatactactt tttcatttaa tctttaaaat tagtgcttca aagctatttg   17460
tggctttctct aaaaatgtgt gcaggagata tcaagggatg aaccatgctt gaactgtcct   17520
taaaaatgga tagaaaccat gttctatatt tcttttgtgt gcatttactt actattccca   17580
gtttctgtgt ccaccccttc tatacccact tgtcatgtat tcccaacaaa cacttgcaca   17640
actccgtggt gtagacagtg ctaattacag tactgatggt gtagaagcaa aagtgcataa   17700
caaccttcag aaacttaaat cttgtttttgt ttgtttgttt tttaaaaagg agatcagcag   17760
gctccagctc cagctggaca ggttgattat acaaaggctt gggaagaata ctacaagaaa   17820
atgggtatgc tctacagttt acgtcttatc cttatctta gtaactaaac ggaagtaaat   17880
acacattaca aagctagaga tgagtaacaa aacatttgct tgtttttaaa aaagaatact   17940
tgaggtaact catccgtatg ttttttataat gttaaagttt atattaaaca agctaggcag   18000
tggtgatacc tgcctttaat cccagtactt ggaaggtaga ggaaggtaga tctccgagtt   18060
tgaggccagc cttgtctaca aagcaagttc caggacactt gggactctta cacagagaaa   18120
ccccatcttg aaaaacaaaa ataataataa acaggaaagg tgacatcctg aaattgactg   18180
gaagttaaga catattaaga gcaattcaag aaattcttat aattgtgctt atggcagaga   18240
aatacaattg tcagcaagtt aataattaaa acctggcatt tccttactag gtaataatat   18300
atggctttat tactgtactt ttgtgtagtg gtctaaatgg aatgagatta aattttaggc   18360
cagcaaagta gttgtatcac tggctttact gcaaaataag cataaaatta gcccacagtg   18420
acattgcttt acatttctaa acatttgcat gcattgaaaa ggaaaaaagg gtgacctttt   18480
aaaaagattt tgacaactat tgaacgtaat tcttaacata accatagtga ttgtgttaat   18540
gtatttcttg tcggtgatac agctttcgac aaagttcctt tcagaacaac tgtttgggga   18600
aatgtgataa caaacttctg cagtgaacag atgggctgat aacttttctt aacctcactg   18660
atcgcaggca ggtctaaca gatggacctt agcactagaa acttgaaacc tgccgggcgt   18720
ggtggcacac acctttaatc ccagcacttg gtgggcagag gcaggcagat ttctgagttc   18780
gaggacagcc tggtctacag agtgagttcc atccaggaca gccagggcta cacagagaaa   18840
ccctgtttgg ggcctggggg gaagaaactt gaaaccttaa aacaatattg tgtgtaatta   18900
gaatgcttac ataatttctt ctctgaactg tgatattttt gagagtagat ggaaatgtat   18960
taatcattcc gtggtagccg tgctgaaata gtgctacttg tttataagaa gctgagcatg   19020
ccaaggtgaa agtgcgctaa ctccattgtt tgttttact gtgtgagtgt tgattgggta   19080
gaaacctggt aactggacta cgtaattaag acagagtagg gtatttgtct gtgctcgatt   19140
atgtcatgtt atttgtttac atgacttcaa ataagtttta tagatatttt gcattctgta   19200
ttcttaaaaa gtatattgac taggacttt gagagtaggt attttaactc aattaggttt   19260
tagttttttaa tattctagca ctttccagca cacagttctg tgaattcctc ttcttctaat   19320
tattattatt attattagag attatttat tttatgtata tgaatgcact gtagctgtac   19380
agatggttgt gagccttcat atggttgttg gaaattgaat tttaagaact ctgctcaatc   19440
cgacccaaag atttatttac tgtcataaat aagtacaccg ttgctgactt cagacatacc   19500
agaagagggc atcatatctc actatgggtg gttgtgagct accatgtggt tgctaggatt   19560
tgaactcagg accttgagaa gagtagtact cttactcgct gagccgtctc ccagcccgaa   19620
ttcctgttct tggttgttat gtctaccgat caatcagttg gcatacatag agacagacac   19680
acacagacag cagcatcgtg caggataact atttggagac tcattttctt catttgtgct   19740
tttttgtaat tatgttgaag atcttaatcc tgtgagcaaa aaaagactaa ctgaacaatt   19800
ccctgtcata tagatcttga aagatactgt tttaggtaat ttaattaatc attgaattag   19860
tatttttaat tatatgatag ttataaaaat catttttccc tgtaaaacat ttaactacaa   19920
ttttttctg aaagtatggt cctgatata tttttagta taatcatgaa ttaagagacc   19980
ttaaagcaaa aggttcagaa tagaagctca ggagatagat gggtaagagt gcttaatgac   20040
ctgggttcaa atcccactac ctatcttaaa agccaatcat gactgttgca tgttgggaag   20100
cagttgcagg agcttgctag ccaactgtgg attcagtaaa aggagcaggg gaataaggta   20160
cagagggata gaagaaggca aaggcaggca gatttctgag ttcgaggcca gcctggtctg   20220
caaagtgagt tccaggacag ccagggctat acagagaaac cctgtctcga aaaaaaagaa   20280
```

-continued

```
aaagaaactc atttgcagtt aacttgcttg tttacaaaat tcattaagta tttctagaat   20340
tcgattttta tcatttgata actcagaatg gtcacttaaa acaaaattac taataaacgc   20400
tgagttttgg gctggtgaga tcgctcagtg ggtaagagca ccctactgct cgtccgaagg   20460
tccggagttc aaatcccagc aaccacatgg tggctcacaa caaccatctg taacaagatc   20520
tgacaccctc ttctggagtg tctgaagaca gctacagtgt acttacatat aataaataaa   20580
taaatcttta aaaaaaaaaa gctagattt acatctttca ttgaacttta tttcagctgc   20640
caggttattt ttagtaaagc tgtgatataa actcagtatg tctttgtgctt cattgtgtgt   20700
gtcataaata tctttcagta cttctctgta gtttaagact agtaatatag tgacccaaat   20760
ggatagaaga gtggtgagct aggtatggtg gcacatatac gcatttagtc ctagtgctca   20820
ggaggcagaa gcaggtagat ctccaaggat gaggtcagcc ctggatagtg agaccctttc   20880
tcagttagca aaaaaaaaaa acaaagcgtt tcttcttttcc ttgtcaaaga ttatatctga   20940
attttcccct ttgtttaaag gtatgtgagt ttgggtcatt ttgtcttttta ctaatcaagc   21000
tgaactaaga gactgaaact tgtgattgaa gtgaaaagag tatctttgcc gtttcttgtg   21060
tgctgcactt tcccccctca tcctccctcc cttgtctaag gcatatttaa gataaacaga   21120
aactgaaggg gcaatgaggc acagataatc ctagcacgtg ggaagttgcc aagaggataa   21180
caagtttttga gccacctgtg ctacataagt gaaaccctgt gttgagataa aaacaagagc   21240
tattgcccct ctgtgccttg cattataagc tgtacttaat ggtacataca tttttgtatt   21300
aaagttgagg gggtgtttta ggtcacaagg gcagacacga ggttattaaa aggcttggaa   21360
gaagtattaa aaaggcaaga tgctgtgttt actagagaag agttactgtg agatttgtaa   21420
ttatagcagc aataaagttc tctgtttccc tgaaggtcaa gcagttcctg ctcctgctgg   21480
ggccccacca ggtggtcagc cggattatag tgcagcctgg gctgagtact atagacagca   21540
agcagcgtat tatgcccaga caagtcccca ggggatgccg cagcatcctc cagcacctca   21600
ggtagaaaac atttgctatt ttttttgttt tgttttttcc actccaacca tgttttctgc   21660
atgtttgtct gtttgtggga atgaatatgt ttttacttggt aa agtataacta   21720
catgaaatat cgtgttgtgt ttggggactc agtttttcta ataatgtttt ctggcatctg   21780
agtgctattt ttgtaatgcc tttgatttta aaaataaatt ttcttccccc cagggatttg   21840
caaatcatgc aagaagccac catcatttat attaaccagt ttttcttttct taaaggattc   21900
actcctgaat tagctccatt taaaggattt tctttaactt tttgtgtatc tcttatgtat   21960
ctcttctgca cagggccaat aataagaagt ggacaataca gtatttgctt cattgtgtgg   22020
gggaaaaaac ctttgttaaa tatatggatg cagacgactt gatgaagatc ttaattttgt   22080
ttttggttta aagtagtgtt tttttccccc ccttttttttt tattttgaaa atgtacaaaa   22140
taactatcac tactgatagg aggttaatat ttctgtgtag aaatgaaaat tggtttgttt   22200
ttagtctttta gtgtagatgt acacattcca gcaaatgtat ttgcaactat tatgtggtcg   22260
atgctttgtg atataaatgt acttttttcaa tgtatacttt cacttttaaa atgcctgttt   22320
tgtgcttttac aataaatgat atgaaacctc ctgtgtcggt aagttggata tgtgggtaaa   22380
ggattcatag tttcttagca gtgataaatt aagatacata tacacgaata tataagcttt   22440
ccccatgaat tactgagttt ttaaacactg gcatgttttt tccccctgttg gagtatagtg   22500
gtagattgga ggttctttttc tgttgtattt ggctatttca gcacaagtaa tcctgatatt   22560
ttcatgtttt tccttctatt tgattaaaaa ctgcatgtgt atacaatgat ctttagtata   22620
cttccattgc attaacagtg acatttcctt ttatacatga gcactatttc agacctgtac   22680
cgctgctaca acagttaacc ttcctgttct tcacttattc ccgagactgt ttcagcataa   22740
ctaattttga acagtttgca gacagtgatt tgaggagttt ataagaaaca ttgtttttttt   22800
catgtaaagc aactctttcc atgtatatat atatattata atagtgtgtt tctctctaaa   22860
ttcaggatag aaaagtaata gaatgtgaaa gtatagctac attgcatctg ccattgaaac   22920
atttgggta tgaaaatgtt caagcttttt ttttctttttt gcagtataga taagctttgt   22980
cttgtaattg cacaagtcca gtcattgaat caaattaatt tttttatgta ctgaatcatt   23040
ttattaatct ttaacattca tgctgaagtt ctgatatttt gttgaaaacc attgttttac   23100
tctgcatatt tgttggctct ttgcatatta atatattaga ctacatgcaa atacagtctg   23160
tcttgccatt gtctgttgaa gtgcaggttt gatccagcca gtatagaact agctctgtag   23220
gggtgaggag gactgtgctg tgtatcatcc ttgattgtgt tccttcaagg agcattgcac   23280
tgtaagtaca tcagagtgac aaaattgatga actgcaacag tgtctttttg tcgatgttcc   23340
acataatgca attgctatat tttgtgtgac tattatgttg gaatacagtg ctgtcatggg   23400
aaaccataac tgcttcttaa catagaataa tacatagttc tgtattttttt ttaagtgagc   23460
ttaatgggta agtattttttg atatgcttta gctaatagct aaagaaaatt gatcagtaac   23520
aaagttgaat agtattatca gtgctcctaa aatgctgttt ttcagtgtaa aatatgccta   23580
tcttttatag tgatatgaaa aacttgaaat gtttaagaca gaagtgcttt tcagtttgca   23640
aagtttaagg acttaacgcc ttttcattaa atgttagttc tatcatctgt ggggagcacc   23700
atttgtatga ggacaaaaat gttgaggtta tagggcagaa aatagtacag ctcattgtgg   23760
atgtttttgaa atgtttttttg attgtttttat gtaacttagt gacttcctt ccccttattt   23820
agatctggat gcatagctct agtatgaagt ttaaagtaa tagttaaggc cttacctgtt   23880
aagatcttaa gctgtagggta ttgacatgaa cagtttcagt attttgcatg atattgttgc   23940
atagatgacc taggaaagtg ttgtggtgca tttagtaatt aaactgaagg aaatagttgg   24000
attcagtatc ataattcaca aattggaggc tgttgatttt gattcgttta aaatttaaaa   24060
tctttattaa ttgcaaacag tgcaattatt tatacttcac agtgccttcc cagaccttcc   24120
accttaggtt ctgctgcaaa aagcaacagg taagcacaac ctaaggccat ctctatataa   24180
atatatcagt acatacatgt tgtccctgtg aggtttgtgg ttgtgtactc actcaagcaa   24240
tctgctgctg ccgctgcccc aaatgtactt tgttatttat ggtaccattc tagtggaaag   24300
tctgttaagt tgttcaagca actgtttaca attttgggtg atgtttttgtt tttggtttttt   24360
ttgggttttt ttttttcccca ttaaaatgag tagattgctg caataactga aaaacatcca   24420
aatctttttttg tgcacccctc ctctcccaga gttatagaaa tgtttaagaa cacttcagta   24480
gttgagacat aggaaatcat ttggtcagaa atttcaacag cctcacaagc ccgtttgtat   24540
gagtaacagg aaattctctt tgccctccaa atctgatctt tttatagttt tattttttatt   24600
atttttcccc tgtattagta ctgctttaaa gtacagtaat tcagtgagat cgtttttggtt   24660
cagtttcata gacagttcta ttttcatgac ttgtatttgc tgacaaggaa gaataaagaa   24720
tactggcagt aagttttggg aaaggcatgt gcatcagtga aatgttactt gtataacaaa   24780
taaccttttca taatctgcat aaccagtagt ttctgattta taatatttta ataactgact   24840
```

-continued

```
gcatttattt ttgccagttt aaaatgtttt gtgttcttag gattgatgtt tagtgcatat  24900
tttgagttaa atgactatct taaagcagca ccatatcagt tgtttttatt cattttctaa  24960
aatgtgctga tcctattaaa aactcctgct tatcttttac aacaaagaaa aatattcaaa  25020
aatactgcct tcattttcac acacagtgct gaagatgctg caagcaccaa atcatagctc  25080
ataaaatcag gtcctgagat agttacccat aaagaggaat cctttgagtg tatgccattg  25140
gtgagccgat gagcatggac catagaaggg ctcaatgtga aaggtaaaat tggcaaatca  25200
taattgagaa atatgactgt gttcccatac ataatatggt ataggggtgta atgtacctgc  25260
tcttgatcac tttcatttta aagtgctatt cacttaaatg ttccatgaat tgtttattgc  25320
accacagttt gtgtgtgtgt gtgtgtgtga gagagagaga gagagagaga gaaataggtc  25380
agatcagtat gttgcaagca gatctttaga gcctgtcaag ttttggttag ttgtagtttc  25440
catttggaaa tgtagatgaa tgcttgtaga tattggagat tgtttctatc ttgtaagaac  25500
ttttcactgg tgctgtaagc atttcaaata gcaccagtct taacctttaa atgggaagta  25560
gaaaaggtga gcccaaagtt tacagatgat tttaatgcta tacatgttag tgtagtgata  25620
cttagaatgc tttgtttgat gtttatttca gaaatgcgta tactagaaaa tcattttaat  25680
attaaaactg gtgacttaat actagttgta aagtgtttca ttaaagaagg atcttggtac  25740
ttaattgata aagtgggttt agataggagt agcaagtgct ctcgatagag aaagttttg   25800
ttcacttcaa tactttgtca ttacaaccag ttcttcctga aaatagttac atgtctagta  25860
aattgatgta gaattaactc gctggagata atctcatttt agttctgcaa attctgcctg  25920
gctttaaaat gcattttcat aatacttaga aataattgta ctgaaaataa ctgctttttt  25980
atttaatcag tcaatcaact tttactacaa atttaattga gggattttt aatttaattg   26040
gtgctttaaa gaagcaaata aatccctggg ttttgttttc ttcagtaaat atcctaaaga  26100
aactctttaa tgtatttgcg agtatatata tattttctta tgcatgctcg atgcattttc  26160
gtcctgagaa aagtgttctc tacagaaact acccgtgtgt taaaagaaga ttggcttaaa  26220
atggctactg tgatgggaac agtgtcttag ggagatgcag cttggacttg aggtaaattg  26280
aatactttac aaactggctt agagtttttgc tttaatgtca ttatatgtaa aagggcacat  26340
gattattgta attttgtatt ctttatggtt tccttaatta aaataataaa tgtacagtga  26400
ttact                                                                26405
```

```
SEQ ID NO: 10           moltype = DNA   length = 4525
FEATURE                 Location/Qualifiers
source                  1..4525
                        mol_type = genomic DNA
                        organism = Mus musculus
SEQUENCE: 10
cctttcttag cagttaaccg agagcggcgc ctccctccgc gaggaaacgg cagcgtgtgc  60
tgtagcgtgg gtatggccga ctactccaca gtgcctccgc cgtcgtctgg ctcggctggc  120
ggcggcggcg gcggagtcgt taacgacgct ttcaaagatg ccctgcagag agcgcgggcag 180
attgcagcaa aaattggggg tgatgctggt acatcattga attcaaatga ctatggttat  240
gggggacaaa aaagacccct agaagatgga gatcagccag atgctaagaa agtacctccc  300
caaaatgact cttttggagc acagttacct ccaatgcatc agcagcaaag atctgtaatg  360
acagaagaat acaaagtccc agatgggatg gtcggattta taattggcag aggaggtgaa  420
cagatctcac gaatacagca ggaatctgga tgcaagatac agatagcacc tgatagtggt  480
ggcctaccag aaaggtcttg tatgctaact ggaacacctg aatctgtcca atcagcaaaa  540
agattattgg accagattgt tgaaaaggga agaccagccc ctggctttca tcatggtgat  600
ggacctggaa atgcagttca ggaaatcatg attccagcca gcaaagcagg actagttatt  660
ggaaaggggg gcgagactat taaacaactt caggaacggg ctggtgttaa aatggtaatg  720
attcaagatg ggcctcaaaa cactggtgct gataaacctc ttaggattac gggtgaccca  780
tacaaagttc agcaagccaa ggaaatggta ttagagttaa ttcgtgatca aggtggtttc  840
agagaagtgc ggaatgagta tggctcaaga ataggaggca atgaagggat agatgtccca  900
attccaagat ttgctgttgg cattgtaata ggaagaaatg gagaaatgat taaaaaaata  960
caaaatgatg ctggtgttcg aattcagttt aagccagatg atggaacaac acctgatagg  1020
atagcacaga taacaggacc tccagacagg tgtcagcacg gtcagaaat aatcacagac  1080
cttctacgaa gtgttcaggc tggcaatcct ggtggaccgg gacctggtgg tcgaggacga  1140
ggtagaggtc aaggaaactg gaatatgggg ccccgggtg gactccagga gtttaatttc  1200
attgtgccaa ctgggaaaac tggactgatc attggaaaag gaggtgaaac cataaaaagc  1260
ataagcaac agtctggtgc aagaatagaa ctgcagagaa gccctccacc taatgcagat  1320
cccaatatga agttatttac aattcggggc actccacagc aaatagacta tgctcgacaa  1380
ctcatagaag agaagattgg gggcccagta aatcctttag ggccacctgt accccatggg  1440
ccccatgggg ttccaggtcc tcatgggcct cctggacctc cagggccggg aactccaatg  1500
ggaccataca accctgcacc ttacaatcca ggaccacctg gcccagctcc tcacggtcct  1560
ccagccccat atgctcccca gggatgggga aatgcgtatc gacattggta gcacaggct   1620
cctcctgacc cagccaaggc aggagcagat ccaaactcgg cagcttgggc tgcttattat  1680
gctcactatt accagcagca ggcacaaccc ccacctgcag ctcctgccgg tgcaccagct  1740
acaacccaaa cgaacggaca aggtaactac ggagatcagc aggctccagc tccagctgga  1800
caggttgatt atacaaaggc ttgggaagaa tactacagaa aaatgggtca agcagttcct  1860
gctcctgctg gggccccacc aggtggtcag ccggattata gtgcagcctg ggctgcgagtac 1920
tatagacagc aagcagcgta ttatgcccag acaagtcccc aggggatgcc gcagcatcct  1980
ccagcacctc agggccaata ataagaagtg gacaatacag tatttgcttc attgtgtgggg 2040
ggaaaaaacc tttgttaaat atatggatgc agacgacttg atgaagatct taattttgtt  2100
tttggttaa agtagtgttt ttttcccccc ctttttttt attttgaaaa tgtacaaaat   2160
aactatcact actgatagga ggtaaatatt tctgtgtaga atgaaaatt ggtttgtttt   2220
tagtctttag tgtagatgta cacattccag caaatgtatt tgcaactatt atgtggtcga  2280
tgctttgtga tataaatgta cttttttcaat gtatactttc acttttaaaa tgcctgtttt  2340
gtgctttaca ataaatgata tgaaacctcc tgtgtcggta agttggatat gtgggtaaag  2400
gattcatagt ttcttagcag tgataaaatta agatacatgt acacgaatat ataagctttc  2460
cccatgaatt actgagtttt taaacactgg catgttttt ccctgttgg agtatagtgg    2520
```

-continued

```
tagattggag gttctttct gttgtatttg gctatttcag cacaagtaat cctgatattt    2580
tcatgtttt cctctattt gattaaaaac tgcatgtgta tacaatgatc tttagtatac    2640
ttccattgca ttaacagtga catttccttt tatacatgag cactatttca gacctgtacc    2700
gctgctacaa cagttaacct tcctgttctt cacttatttc cgagactgtt tcagcataac    2760
taattttgaa cagtttgcag acagtgattt gaggagttta taagaaacat tgttttttc    2820
atgtaaagca actctttcca tgtatatata tatattataa tagtgtgtt ctctctaaat    2880
tcaggataga aaagtaatag aatgtgaaag tatagctaca ttgcatctgc cattgaaaca    2940
tttgggtat gaaaatgttc aagcttttt tttcttttg cagtatagat aagctttgtc    3000
ttgtaattgc acaagtccag tcattgaatc aaattaattt ttttatgtac tgaatcattt    3060
tattaatctt taacattcat gctgaagttc tgatattttg ttgaaaacca ttgttttact    3120
ctgcatattt gttggctctt tgcatattaa tatattagac tacatgcaaa tacagtctgt    3180
cttgccattg tctgttgaag tgcaggtttg atccagccag tatagaacta gctctgtagg    3240
ggtgaggagg actgtgctgt gtatcatcct tgattgtgtt ccttcaagga gcattgcact    3300
gtaagtacat cagagtgaca aattgatgaa ctgcaacagt gtcttttgt cgatgttcca    3360
cataatgcaa ttgctatatt ttgtgtgact attatgttgg aatacagtgc tgtcatggga    3420
aaccataact gcttcttaac atagaataat acatagttct gtatttttt taagtgagct    3480
taatgggtaa gtattttga tatgctttag ctaaatagcta aagaaaattg atcagtaaca    3540
aagttgaata gtattatcag tgctcctaaa atgctgtttt tcagtgtaaa atatgccat    3600
cttttatagt gatatgaaaa acttgaaatg tttaagacag aagtgctttt cagtttgcaa    3660
agtttaagga cttaacgcct tttcattaaa tgttagttct atcatctgtg gggagcacca    3720
tttgtatgag gacaaaaatg ttgaggttat agggcagaaa atagtacagc tcattgtgga    3780
tgttttgaaa tgtttttga ttgttttatg taacttagtg acttcccttc cccttattta    3840
gatctggatg catagctcta gtatgaagtt taatagtaat agttaaggcc ttacctgtta    3900
agatcttaag tgtagggtat tgacatgaac agtttcagta ttttgcatga tattgttgca    3960
tagatgacct aggaaagtgt tgtggtgcat ttagtaatta aactgaagga aatagttgga    4020
ttcagtatca taattcacaa attggaggct gttgattttg attcgtttaa aatttaaaat    4080
ctttattaat tgcaaacagt gcaattattt atacttcaca gtgccttccc agaccttcca    4140
ccttaggttc tgctgcaaaa agcaacagtg ctgaagatgc tgcaagcacc aaatcatagc    4200
tcataaaatc aggtcctgag atagttaccc ataaagagga atcctttgag tgtatgccat    4260
tggtgagccg atgagcatgg accatagaag ggctcaatgt agaagaaact acccgtgtgt    4320
taaaagaaga ttggcttaaa atggctactg tgatgggaac agtgtcttag ggagatgcag    4380
cttggacttg aggtaaattg aatactttac aaactggctt agagttttgc tttaatgtca    4440
ttatatgtaa aagggcacat gattattgta attttgtatt ctttatggtt tccttaatta    4500
aaataataaa tgtacagtga ttact                                         4525
```

```
SEQ ID NO: 11          moltype = DNA  length = 800
FEATURE                Location/Qualifiers
source                 1..800
                       mol_type = genomic DNA
                       organism = Mus musculus
SEQUENCE: 11
cagggatggg gaaatgcgta tccacattgg cagcaacagg ctcctcctga cccagccaag    60
gcaggagcag atccaaactc ggcagcttgg gctgcttatt atgctcacta ttaccagcag    120
caggcacaac ccccacctgc agctcctgcc ggtgcaccag ctacaaccca acgaacggga    180
caaggagatc agcaggctcc agctccagct ggacaggttg attatacaaa ggcttgggaa    240
gaatactaca agaaaatggg tcaagcagtt cctgctcctg ctggggcccc accaggtggt    300
cagccggatt atagtgcagc ctgggctgag tactatagac agcaagcagc gtattatgcc    360
cagacaagtc cccaggggat gccgcagcat cctccagcac ctcaggggca ataataagaa    420
gtggacaata cagtatttgc ttcattgtgt ggggggaaaa acctttgtta aatatatgga    480
tgcagacgac ttgatgaaga tcttaatttt gttttttggtt taaagtagtg ttttttttccc    540
cccctttttt tttattttga aaatgtacaa aataactatc actactgata ggaggttaat    600
atttgtgt agaaatgaaa attggtttgt ttttagtctt tagtgtatgt gtacacattc    660
cagcaaatgt atttgcaact attatgtggt cgatgctttg tgatataaat gtactttttc    720
aatgtatact ttcactttta aaatgcctgt tttgtgcttt acaataaatg atatgaaacc    780
tcctgtgtcg gtaagttgga                                               800
```

```
SEQ ID NO: 12          moltype = DNA  length = 2526
FEATURE                Location/Qualifiers
source                 1..2526
                       mol_type = genomic DNA
                       organism = Mus musculus
SEQUENCE: 12
gcctcgacgg ccgccatctt cttcctttct tagcagttaa ccgagagcgg cgcctccctc    60
cgcgaggaaa cggcagcgtg tgctgtagcg tgggtatggc cgactactcc acagtgcctc    120
cgccgtcgtc tggctcggct ggcggcggcg gcggcggagt cgttaacgac gctttcaaag    180
atgccctgca gagagcgcgg cagattgcag caaaaattgg gggtgatgct ggtacatcat    240
tgaattcaaa tgactatggt tatggggggac aaaaaagacc cttagaagat ggagatcagc    300
cagatgctaa gaaagtacct ccccaaaatg actcttttgg agcacagtta cctccaatgc    360
atcagcagca aagcagatct gtaatgacag aagaatacaa agtcccagat gggatggtcg    420
gatttataat tggcagagga ggtgaacaga tctcacgaat acagcaggaa tctggatgca    480
agatacagat agcacctgat agtggtggcc taccagaaag gtcttgtatg ctaactggaa    540
cacctgaatc tgtccaatca gcaaaaagat tattggacca gattgttgaa aagggaagac    600
cagccctgg ctttcatcat ggtgatgac ctgaaatgc agttcaggaa atcatgattc    660
cagccagcaa agcaggacta gttattggaa agggggggcga gactattaaa caacttcagg    720
aacgggctgc tgttaaaatg gtaatgattc aagatgggcc tcaaaacact ggtgctgata    780
aacctcttag gattacgggt gacccataca aagttcagca agccaaggaa atggtattag    840
```

```
agttaattcg tgatcaaggt ggtttcagag aagtgcggaa tgagtatggc tcaagaatag   900
gaggcaatga agggatagat gtcccaattc caagatttgc tgttggcatt gtaataggaa   960
gaaatggaga aatgattaaa aaaatacaaa atgatgctgg tgttcgaatt cagtttaagc  1020
cagatgatgg aacaacacct gataggatag cacagataac aggacctcca gacaggtgtc  1080
agcacgctgc agaaataatc acagaccttc tacgaagtgt tcaggctggc aatcctggtg  1140
gaccgggacc tggtggtcga ggacgaggta gaggtcaaga aaactggaat atggggccc   1200
cgggtggact ccaggagttt aatttcattg tgccaactgg gaaaactgga ctgatcattg  1260
gaaaaggagg tgaaaccata aaaagcataa gccaacagtc tggtgcaaga atagaactgc  1320
agagaagccc tccacctaat gcagatccca atatgaagtt atttacaatt cggggcactc  1380
cacagcaaat agactatgct cgacaactca tagaagagaa gattgggggc ccagtaaatc  1440
ctttagggcc acctgtaccc catgggcccc atggggttcc aggtcctcat gggcctcctg  1500
gacctccagg gccgggaact ccaatgggac catacaaccc tgcaccttac aatccaggac  1560
cacctggccc agctcctcac ggtcctccag ccccatatgc tccccaggga tggggaaatg  1620
cgtatccaca ttggcagcaa caggctcctc ctgacccagc caaggcagga gcagatccaa  1680
actcggccagc ttgggctgct tattatgtc actattacca gcagcaggca caaccccac  1740
ctgcagctcc tgccggtgca ccagctacaa cccaaacgaa cggacaagga gatcagcagg  1800
ctccagctcc agctggacag gttgattata caaaggcttg ggaagaatac tacaagaaaa  1860
tgggtcaagc agttcctgct cctgctgggg ccccaccagg tggtcagccg gattatagtg  1920
cagcctgggc tgagtactat agacagcaag cagcgtatta tgcccagaca agtccccagg  1980
ggatgccgca gcatcctcca gcacctcagg gatttgcaaa tcatgcaaga agccaccatc  2040
atttatatta accagttttt ctttcttaaa ggattcactc ctgaattagc tccatttaaa  2100
ggattttctt taactttttg tgtatctctt atgtatctct tctgcacagg gccaataata  2160
agaagtggac aatacagtat ttgcttcatt gtgtgggggga aaaacctttt gttaaatata  2220
tggatgcaga cgacttgatg aagatcttaa ttttgttttt ggtttaaagt agtgtttttt  2280
tccccccctt ttttttttatt ttgaaaatgt acaaaataac tatcactact gataggaggt  2340
taatatttct gtgtagaaat gaaaattggt ttgtttttag tctttagtgt agatgtacac  2400
attccagcaa atgtatttgc aactattatg tggtcgatgc tttgtgatat aaatgtactt  2460
tttcaatgta tactttcact tttaaaatgc ctgttttgtg ctttacaata aatgatatga  2520
aacctc                                                             2526


SEQ ID NO: 13         moltype = DNA  length = 809
FEATURE               Location/Qualifiers
source                1..809
                      mol_type = genomic DNA
                      organism = Mus musculus
SEQUENCE: 13
gaaatgcgta tccacattgg cagcaacagg ctcctcctga cccagccaag gcaggagcag   60
atccaaactc ggcagcttgg gctgcttatt atgctcacta ttaccagcag caggcacaac  120
ccccacctgc agctcctgcc ggtgcaccag ctacaaccca aacgaacgga caaggagatc  180
agcaggctcc agctccagct ggacaggttg attatacaaa ggcttgggag gaatactaca  240
agaaaatggg tcaagcagtt cctgctcctg ctggggcccc accaggtggt cagccggatt  300
atagtgcagc ctgggctgag tactatagac agcaagcagc gtattatgcc cagacaagtc  360
cccaggggat gccgcagcat cctccagcac ctcagtgcct tcccagacct tccaccttag  420
gttctgctgc aaaaagcaac agtgctgaag atgctgaag caccaaatca tagctcataa  480
aatcaggtcc tgagatagtt acccataaag aggaatcctt tgagtgtatg ccattggtga  540
gccgatgagc atggaccata gaagggctca atgtagaagg taaaattggc aaatcataat  600
tgagaaatat gactgtgttc ccatacataa tatggtatag ggtgtaatgt acctgctctt  660
gatcactttc attttaaagt gctattcact taaatgttcc atgaattgtt tattgcacca  720
cagtttgtgt gtgtgtgtgt gtgtgagaga gagagagaga gagagaaa taggtcagat  780
cagtatgttg caagcagatc tttagagcc                                    809


SEQ ID NO: 14         moltype = DNA  length = 1040
FEATURE               Location/Qualifiers
source                1..1040
                      mol_type = genomic DNA
                      organism = Mus musculus
SEQUENCE: 14
gatggtcgtg caagaatgtg atagagcctc gacggccgcc atcttcttcc tttcttagca   60
gttaaccgag agcggcgcct ccctccgcga ggaaacggca gcgtgtgctg tagcgtgggt  120
atggccgact actccacagt gcctccgccg tcgtctggct cggctggcgg cggcggcggc  180
ggagtcgtta acgacgcttc caaagatgcc tgcagaaga cgcggcagat tgcagcaaaa  240
attgggggtg atgctggtac atcattgaat tcaaatgact atggttatgg gggacaaaaa  300
agaccctag aagatggaga tggctcttgg acaaatccga gcagtaccac acactgggag  360
ggaatgccct ctccttttaa agatcagcca gatgctaaga aagtacctcc ccaaaatgac  420
tcttttggag cacagttacc tccaatgcat cagcagcaaa gatctgtaat gacagaagaa  480
tacaaagtcc cagatgggat ggtcggattt ataattggca gaggaggtga acagatccta  540
cgaatacagc aggaatctgg atgcaagata cagatagcac ctgatagtgg tggcctacca  600
gaaaggtctt gtatgctaac tggaacacct gaatctgtcc aatcagcaaa aagattattg  660
gaccagattg ttgaaaaggg aagaccagcc cctggctttc atcatggtga tggacctgga  720
aatgcagttc aggaaatcat gattccagcc agcaaagcag gactagttat tggaaagggg  780
ggcagacta ttaaacaact tcaggaacgg gctggtgtta aatggtaat gattcaagat  840
gggcctcaaa acactggtgc tgataaacct cttaggatta cgggtgaccc atacaaagtt  900
cagcaagcca aggaaatggt attagagtta attcgtgatc aaggtggttt cagagaagtg  960
cggaatgagt atggctcaag aataggaggc aatgaaggga tagatgtccc aattccaaga  1020
tttgctgttg gcattgtaat                                              1040
```

-continued

```
SEQ ID NO: 15            moltype = DNA   length = 796
FEATURE                 Location/Qualifiers
source                  1..796
                        mol_type = genomic DNA
                        organism = Mus musculus
SEQUENCE: 15
cccatatgct cccccagggat ggggaaatgc gtatccacat tggcagcaac aggctcctcc   60
tgacccagcc aaggcaggag cagatccaaa ctcggcagct tgggctgctt attatgctca  120
ctattaccag cagcaggcac aaccccacc tgcagctcct gccggtgcac cagctacaac  180
ccaaacgaac ggacaaggag atcagcaggc tccagctcca gctggacagg ttgattatac  240
aaaggcttgg gaagaatact acaagaaaat gggtcaagca gttcctgctc ctgctggggc  300
cccaccaggt ggtcagccgg attatagtgc agcctgggct gagtactata gacagcaagc  360
agcgtattat gcccagacaa gtccccaggg gatgccgcag catcctccag cacctcaggg  420
ccaataataa gaagtggaca atacagtatt tgcttcattg tgtgggggaa aaaacctttg  480
ttaaatatat ggatgcagac gacttgatga agatcttaat tttgttttg gtttaaagta  540
gtgtttttt cccccccttt ttttttattt tgaaaatgta caaaataact atcactactg  600
ataggaggtt aatatttctg tgtagaaatg aaaattggtt tgtttttagt ctttagtgta  660
gatgtacaca ttccagcaaa tgtatttgca actattatgt ggtcgatgct ttgtgatata  720
aatgtacttt ttcaatgtat actttcactt ttaaaatgcc tgtttgtgc tttacaataa  780
atgatatgaa acctcc                                                796

SEQ ID NO: 16            moltype = DNA   length = 585
FEATURE                 Location/Qualifiers
source                  1..585
                        mol_type = genomic DNA
                        organism = Mus musculus
SEQUENCE: 16
aatttcattg tgccaactgg gaaaactgga ctgatcattg gaaaaggagg tgaaaccata   60
aaaagcataa gccaacagtc tggtgcaaga atagaactgc agagaagccc tccacctaat  120
gcagatccca atatgaagtt atttacaatt cggggcactc cacagcaaat agactatgct  180
cgacaactca tagaaagaa gattgggggc ccagtaaatc ctttaggggc acctgtaccc  240
catgggcccc atgggggttcc aggtcctcat gggcctcctg gacctccagg gccgggaact  300
ccaatgggac catacaaccc tgcaccttac aatccaggac cacctggccc agctcctcac  360
ggtcctccag ccccatatgc tccccaggga tggggaaatg cgtatccaca ttggcagcaa  420
caggctcctc ctgacccaga aacaacttga gccgggcggc agtggtgaac gccattagtc  480
ccagcacaca ggagacagaa gcaggcagat ctctgagttc gaggacagcc tgatctacag  540
agcatgttca aggatagcca gggacacaca ggaaaaatct tatct               585

SEQ ID NO: 17            moltype = DNA   length = 2374
FEATURE                 Location/Qualifiers
source                  1..2374
                        mol_type = genomic DNA
                        organism = Mus musculus
SEQUENCE: 17
ttcctttctt agcagttaac cgagagcggc gcctccctcc gcgaggaaac ggcagcgtgt   60
gctgtagcgt gggtatggcc gactactcca cagtgcctcc gccgtcgtct ggctcggctg  120
gcggcggcgg cggcggagtc gttaacgacg ctttcaaaga tgccctgcag agagcgcggc  180
agattgcagc aaaaattggg ggtgatgctg gtacatcatt gaattcaaat gactatggtt  240
atgggggaca aaaaagaccc ttagaagatg gagatcagcc agatgctaag aaagtacctc  300
cccaaaatga ctcttttgga gcacagttac ctccaatgca tcagcagcaa agcagatctg  360
taatgacaga agaatacaaa gtcccagatg ggatggtcgg attttataatt ggcagaggag  420
gtgaacagat ctcacgaata cagcagggat ctggatgcaa gatacagata gcacctgata  480
gtggtggcct accagaaagg tcttgtatgc taactggaac acctgaatct gtccaatcag  540
caaaaagatt attggaccag attgttgaaa agggaagacc agccctggc tttcatcatg  600
gtgatggacc tggaaatgca gttcaggaaa tcatgattcc agccagcaaa gcaggactag  660
ttattggaaa gggggggcgag actattaaac aacttcagga acgggctggt gttaaaatgg  720
taatgattca agatgggcct caaaacactg gtgctgataa acctcttagg attacgggtg  780
acccatacaa agttcagcaa gccaaggaaa tggtattaga gttaattcgt gatcaaggtg  840
gtttcagaga agtgcggaat gagtatggct caagaatagg aggcaatgaa gggatagatg  900
tcccaattcc aagatttgct gttggcattg taataggaag aaatggagaa atgattaaaa  960
aaatacaaaa tgatgctggt gttcgaattc agtttaagcc agatgga acaacacctg 1020
ataggatagc acagataaca ggacctccag acaggtgtca gcacgctgca gaaataatca 1080
cagaccttct acgaagtgtt caggctggca atcctggtgg accgggacct ggtggtcgag 1140
gacgaggtag aggtcaagga aactggaata tggggcccc gggtgactc caggagttta 1200
atttcattgt gccaactggg aaaactggac tgatcattg aaaaggaggt gaaaccataa 1260
aaagcataag ccaacagtct ggtgcaagaa tagaactgc agagaagccct ccacctaatg 1320
cagatcccaa tatgaagtta tttacaattc ggggcactcc acagcaaata gactatgctc 1380
gacaactcat agaagagaag attgggggcc cagtaaatcc tttaggggcca cctgtacccc 1440
atgggcccca tggggttcca ggtcctcatg gcctcctgg acctccaggg ccgggaactc 1500
caatgggacc atacaaccct gcaccttaca atccaggacc acctggccca gctcctcacg 1560
gtcctccagc cccatatgct ccccagggat gggggaaatg cgtatccaca ttggcagcaac 1620
aggctcctc tgacccagcc aaggcaggag cagatccaaa ctcggcagct tgggctgctt 1680
attatgctca ctattaccag cagcaggcac aaccccacc tgcagctcct gccggtgcac 1740
cagctacaac ccaaacgaac ggacaaggta actacgagta tcagcaggct ccagctccag 1800
ctggacaggt tgattataca aaggcttggg aagaatacta caagaaaatg ggtcaagcag 1860
ttcctgctcc tgctggggcc ccaccaggtg gtcagccgga ttatagtgca gcctgggctg 1920
```

-continued

```
agtactatag acagcaagca gcgtattatg cccagacaag tccccagggg atgccgcagc  1980
atcctccagc acctcagggc caataataag aagtggacaa tacagtattt gcttcattgt  2040
gtggggaaaa aaacctttgt taaatatatg gatgcagacg acttgatgaa gatcttaatt  2100
ttgtttttgg tttaaagtag tgtttttttc ccccccttt tttttatttt gaaaatgtac  2160
aaaataacta tcactactga taggaggtta atatttctgt gtagaaatga aaattggttt  2220
gtttttagtc tttagtgtag atgtacacat tccagcaaat gtatttgcaa ctattatgtg  2280
gtcgatgctt tgtgatataa atgtacttt tcaatgtata cttttcacttt taaaatgcct  2340
gttttgtgct ttacaataaa tgatatgaaa cctc  2374
```

SEQ ID NO: 18   moltype = DNA   length = 3163
FEATURE   Location/Qualifiers
source   1..3163
                           mol_type = genomic DNA
                           organism = Mus musculus
SEQUENCE: 18

```
gatagagcct cgacggccgc catcttcttc ctttcttagc agttaaccga gagcggcgcc  60
tccctccgcg aggaaacggc agcgtgtgct gtagcgtggg tatggccgac tactccacag  120
tgcctccgcc gtcgtctggc tcggctggcg gcggcggcgg cggagtcgtt aacgacgctt  180
tcaaagatgc cctgcagaga gcgcgggcaga ttgcagcaaa aattgggggt gatgctggta  240
catcattgaa ttcaaatgac tatggttatg ggggacaaaa aagaccctta gaagatggag  300
atcagccaga tgctaagaaa gtacctcccc aaaatgactc ttttggagca cagttacctc  360
caatgcatca gcagcaaaga tctgtaatga cagaagaata caaagtccca gatgggatgg  420
tcggatttat aattggcaga ggaggtgaac agatctcagc aatacagcag gaatctggat  480
gcaagataca gatagcacct gatagtggtg gcctaccaga aaggtcttgt atgctaactg  540
gaacacctga atctgtccaa tcagcaaaaa gattattgga ccagattgtt gaaaagggaa  600
gaccagcccc tggctttcat catggtgatg gacctggaaa tgcagttcag gaaatcatga  660
ttccagccag caaagcagga ctagttattg gaaagggggg cgagactatt aaacaacttc  720
aggaacgggc tggtgttaaa atggtaatga ttcaagatgg gcctcaaaac actggtgctg  780
ataaacctct taggattacg ggtgacccat acaaagttca gcaagccaag gaaatggtat  840
tagagttaat tcgtgatcaa ggtggtttca gagaagtgcg gaatgagtat ggctcaagaa  900
taggaggcaa tgaagggata gatgtcccaa ttccaagatt tgctgttcga attgtaatag  960
gaagaaatgg agaaatgatt aaaaaaatac aaaatgatgc tggtgttcga attcagttta  1020
agccagatga tggaacaaca cctgatagga tagcacagat aacaggacct ccagacaggt  1080
gtcagcacgc tgcagaaata atcacagacc ttctacgaag tgttcaggct ggcaatcctg  1140
gtggaccggg acctggtggt cgaggacgag gtagaggtca aggaaactgg aatatgggc  1200
ccccgaggag actccaggag tttaatttca ttgtgccaac tgggaaaact ggactgattca  1260
ttggaaaagg aggtgaaacc ataaaaagca taagccaaca gtctggtgca agaatagaac  1320
tgcagagaag ccctccacct aatgcagatc ccaatatgaa gttatttaca attcggggca  1380
ctccacagca aatagactat gctcgacaac tcatagaaga gaagattggg ggcccagtaa  1440
atcctttagg gccacctgta ccccatgggc cccatgggt tccaggtcct catgggcctc  1500
ctggacctcc agggccggga actccaatgg gaccatacaa ccctgcacct tacaatccag  1560
gaccacctgg cccagctcct cacggtcctc cagccccata tgctccccag ggatggggaa  1620
atgcgtatcc acattggcag caacaggctc ctcctgaccc aggtaaaggg taatctacta  1680
ttaatatgtt agcttcattg tatactgctg gtcacagtcc taaactacct actgcatgtg  1740
cgatctttta aggcacttcc cctgcagtat aggatgacac tgctgtctag tatgttgtga  1800
ctttattcat agtagttcat atttataaaa taatgtgttt aacactacta tgttctcaaa  1860
taataagtat aaattatttt gttgctgata aaattaaggt cccttaatta ccaaaacagg  1920
agagtacatt tattttctc tcttataatc tggatttttt atagcatggt gttctcatta  1980
tatcttccca tactggtttt gtttctactc catcatgatc aggaatcaac ttggaaatag  2040
gacttttttc tttgttttc ttttgtttt tcactttgca gccctgtag tttgattcat  2100
ttgaaaagtt tgagagaacg tgttctgagt tatagatttc tacatagtgt attttttaca  2160
aatcaaagtt gatttagatg taattgcttt gttggatttg ttaaatgttt taatattta  2220
atgtatgtta aattcatagg aacttttctc agtcaccaaa taaccacaac ttctattcat  2280
ttaagaaaca acttgagccg ggcggcagtg tgaacgcca ttagtcccag cacacaggag  2340
acagaagcag gcagatctct gagttcgagg acagcctgat ctacagagca tgttcaagga  2400
tagccaggga cacacaggaa aaatcctatc tcaaaaaaga gagagagaga aaaaaaaaga  2460
aaaagaaata acttgaattg tgatggcatc tgaatatgaa agatgagtgt cgattagcaa  2520
aagagtgcat tccagggca gacaagaagc atcagttatg acttccttg taagggtttt  2580
aaaagtgaat cttctccttc taacaagcat ttagacgtgg catgatcaaa tagagttgga  2640
acaggttata ttgaaggttt tatatggcta attcaatttc ataggattgc tcaaatgagt  2700
gatgtttgat attagcactg taaggtctgg tcagtgaata tgcttccag cattacctgg  2760
aagactgact cggcaatatg ttgcttaaat ttgacttgga agtatacagt attttattgt  2820
ctgcattttc aagaaactgc tgctgtgagg ttgagattta tatagatgtc aagaatttaa  2880
tcttcaatct taaatggata atatgcgaat tcatcctatg taaaagaata gtgagctctt  2940
agtacctgct agttgcaagt cacaaaaaaa aagtttgggt tttggtttgg tttttgagtt  3000
acaaaggtga gcttagtgca gaggatgctg gttgtggtga cttgggaata gggcagcagc  3060
taaaacttga gtattggggg atgctaaaag aattaataga aaggataagc cacagtgctg  3120
gggataagca ttagaacctg agttctatcc ctcccacctg ttg  3163
```

SEQ ID NO: 19   moltype = DNA   length = 6523
FEATURE   Location/Qualifiers
source   1..6523
                           mol_type = genomic DNA
                           organism = Mus musculus
SEQUENCE: 19

```
gctgaaacat cgcgggattt ccgtctgcca gtctccgccc ctttacggca cgatggtcgt  60
```

-continued

```
gcaagaatgt gatagagcct cgacggccgc catcttcttc ctttcttagc agttaaccga    120
gagcggcgcc tccctccgcg aggaaacggc agcgtgtgct gtagcgtggg tatggccgac    180
tactccacag tgcctccgcc gtcgtctggc tcggctggcg gcggcggcgg cggagtcgtt    240
aacgacgctt tcaaagatgc cctgcagaga gcgcggcaga ttgcagcaaa aattgggggt    300
gatgctggta catcattgaa ttcaaatgac tatggttatg ggggacaaaa aagacccatt    360
gaagatggag atcagccaga tgctaagaaa gtacctcccc aaaatgactc ttttggagca    420
cagttacctc caatgcatca gcagcaaaga tctgtaatga cagaagaata caaagtccca    480
gatgggatgg tcggatttat aattggcaga ggaggtgaac agatctcacg aatacagcag    540
gaatctggat gcaagataca gatagcacct gatagtggtg gcctaccaga aaggtcttgt    600
atgctaactg gaacacctga atctgtccaa tcagcaaaaa gattattgga ccagattgtt    660
gaaaagggaa gaccagcccc tggctttcat catggtgatg gacctggaaa tgcagttcag    720
gaaatcatga ttccagccag caaagcagga ctagttattg gaaagggggg cgagactatt    780
aaacaacttc aggaacgggc tggtgttaaa atggtaatga ttcaagatgg gcctcaaaac    840
actggtgctg ataaacctct taggattacg ggtgacccat acaaagttca gcaagccaag    900
gaaatggtat tagagttaat tcgtgatcaa ggtggtttca gagaagtgcg gaatgagtat    960
ggctcaagaa taggaggcaa tgaagggata gatgtcccaa ttccaagatt tgctgttggc   1020
attgtaatag gaagaaatgg agaaatgatt aaaaaaatac aaaatgatgc tggtgttcga   1080
attcagttta agcagatgaa tggaacaaca cctgatagga tagcacagat aacaggacct   1140
ccagacaggt gtcagcacgc tgcagaaata atcacagacc ttctacgagg tgttcaggct   1200
ggcaatcctg gtggaccggg acctggtggt cgaggacgag gtagaggtca aggaaactgg   1260
aatatggggc ccccgggtgg actccaggag tttaatttca ttgtgccaac tgggaaaact   1320
ggactgatca ttgaaaaagg aggtgaaacc ataaaaagca taagccaaca gtctggtgca   1380
agaatagaac tgcagagaag ccctccacct aatgcagatc ccaatatgaa gttatttaca   1440
attcggggca ctccacagca aatagactat gctcgacaac tcatagaaga gaagattggg   1500
ggcccagtaa atcctttagg gccacctgta ccccatgggc cccatggggt tccaggtcct   1560
catgggcctc ctggacctcc agggccggga actccaatgg gaccatacaa ccctgcacct   1620
tacaatccag gaccacctgg cccagctcct cacggtcctc cagccccata tgctccccag   1680
ggatggggaa atgcgtatcc acattggcag caacaggctc ctcctgaccc agccaaggca   1740
ggagcagatc caaactcggc agcttgggct gcttattatg ctcactatta ccagcagcag   1800
gcacaacccc cacctgcagc tcctgccggt gcaccagcta caacccaaac gaacggacaa   1860
ggtaactacg gagatcagca ggctccagct ccagctgcac aggttgatta tacaaaggct   1920
tgggaagaat actacaagaa aatgggtcaa gcagttcctg ctcctgctgg ggccccacca   1980
ggtggtcagc cggattatag tgcagcctgg gctgagtact atagacagca agcagcgtat   2040
tatgcccaga caagtcccca ggggatgccg cagcatcctc cagcacctca gggccaataa   2100
taagaagtgg acaatacagt atttgcttca ttgtgtgggg gaaaaaacct ttgttaaata   2160
tatggatgca gacgacttga tgaagatctt aattttgttt ttggtttaaa gtagtgtttt   2220
tttccccccc tttttttta ttttgaaaat gtacaaaata actatcacta ctgatagga   2280
gttaatattt ctgtgtagaa atgaaaattg gtttgttttt agtctttagt gtagatgtac   2340
acattccagc aaatgtattt gcaactatta tgtggtcgat gctttgtgat ataaatgtac   2400
tttttcaatg tatactttca cttttaaaat gcctgttttg tgctttacaa taaatgataa   2460
gaaacctcct gtgtcggtaa gttggatatg tgggtaaagg attcatagtt tcttagcagt   2520
gataaattaa gatacatgta cacgaatata taagctttcc ccatgaatta ctgagttttt   2580
aaacactggc atgtttttc ccctgttgga gtatagtggt agattggagg ttcttttctg    2640
ttgtatttgg ctatttcagc acaagtaatc ctgatatttt catgtttttc cttctatttg   2700
attaaaaact gcatgtgtat acaatgatct ttagtatact tccattgcat taacagtgac   2760
atttcctttt atacatgagc actatttcag acctgtaccg ctgctacaac agttaacctt   2820
cctgttcttc acttatttcc gagactgttt cagcataact aattttgaac agtttgcaga   2880
cagtgatttg aggagtttat aagaaacatt gttttttttca tgtaaagcaa ctctttcag   2940
gtatatatat atattataat agtgtgtttc tctctaaatt caggatagaa aagtaataga   3000
atgtgaaagt atagctacat tgcatctgcc attgaaacat ttgggggtatg aaaatgttca   3060
agcttttttt ttcttttttgc agtatagata agctttgtct tgtaattgca caagtccagt   3120
cattgaatca aattaatttt tttatgtact gaatcattt attaatcttt aacattcatg    3180
ctgaagttct gatattttgt tgaaaaccat tgttttactc tgcatatttg ttggctcttt   3240
gcatattaat atattagact acatgcaaat acagtctgtc ttgccattgt ctgttgaagt   3300
gcaggtttga tccagccagt atagaactag ctctgtaggg gtgaggagga ctgtgctgtg   3360
tatcatcctt gattgtgttc cttcaaggag cattgcactg taagtacatc agagtgacaa   3420
attgatgaac tgcaacagtg tcttttttgtc gatgttccac ataatgcaat tgctatattt   3480
tgtgtgacta ttatgttgga atacagtgct gtcatgggaa accataactg cttcttaaca   3540
tagaataata catagttctg tattttttttt aagtgagctt aatgggtaag tatttttgat   3600
atgctttagc taatagctaa agaaaattga tcagtaacaa agttgaatag tattatcagt   3660
gctcctaaaa tgctgttttt cagtgtaaaa tatgcctatc ttttatgtg atatgaaaaa    3720
cttgaaatgt ttaagacaga agtgcttttc agtttgcaaa gtttaaggac ttaacgcctt   3780
ttcattaaat gttagttcta tcatctgtgg ggagcaccat ttgtatgagg acaaaaatgt   3840
tgaggttata gggcagaaaa tagtacagct cattgtggat gtttttgaaat gtttttttgat   3900
tgtttttatgt aacttagtga cttcccttcc ccttatttga atctggatgc atagctctag   3960
tatgaagttt aatagtaata gttaaggcct tacctgttaa gatcttaagt gtagggtatt   4020
gacatgaaca gtttcagtat tttgcatgat attgttgcat agatgaccta ggaaagtgtt   4080
gtggtgcatt tagtaattaa actgaaggaa atagttggat tcagtatcat aattcacaaa   4140
ttggaggctg ttgattttga ttcgtttaaa atttaaaatc tttattaatt gcaaacagtg   4200
caattatttta tacttcacag tgccttccca gaccttccac cttaggttct gctgcaaaaa   4260
gcaacaggta gcacacaacct aaggccatct ctatataaat atcagtac atacatgttg     4320
tccctgtgag gtttgtggtt gtgtactcac tcaagcaatc tgctgctgcc gctgccccaa    4380
atgtactttg ttatttatgg taccattcta gtggaaagtc tgttaagttg ttcaagcaac    4440
tgtttacaat tttgggtgat gttttgtttt tggttttttt gggtttttttt tttccccatt   4500
aaaatgagta gattgctgca ataactgaaa aacatccaaa tcttttttgtg cacccctcct   4560
ctcccagagt tatagaaatg tttaagaaca cttcagtagt tgagacatag gaaatcattt    4620
```

-continued

```
ggtcagaaat ttcaacagcc ttcacagccc gtttgtatga gtaacaggaa attctctttg  4680
ccctccaaat ctgatctttt tatagtttta tttttattat ttttcccctg tattagtact  4740
gctttaaagt acagtaattc agtgagatcg ttttggttca gtttcataga cagttctatt  4800
ttcatgactt gtatttgctg acaaggaaga ataaagaata ctggcagtaa gttttgggaa  4860
aggcatgtgc atcagtgaaa tgttacttgt ataacaaata acctttcata atctgcataa  4920
ccagtagttt ctgatttata atattttaat aactgactgc atttattttt gccagtttaa  4980
aatgttttgt gttcttagga ttgatgttta gtgcatattt tgagttaaat gactatctta  5040
aagcagcacc atatcagttg tttttattca ttttctaaaa tgtgctgatc ctattaaaaa  5100
ctcctgctta tcttttacaa caaagaaaaa tattcaaaaa tactgccttc attttcacac  5160
acagtgctga agatgctgca agcaccaaat catagctcat aaaatcaggt cctgagatag  5220
ttacccataa agaggaatcc tttgagtgta tgccattggt gagccgatga gcatggacca  5280
tagaagggct caatgtagaa ggtaaaattg gcaaatcata attgagaaat atgactgtgt  5340
tcccatacat aaatatggtat agggtgtaat gtacctgctc ttgatcactt tcattttaaa  5400
gtgctattca cttaaatgtt ccatgaattg tttattgcac cacagtttgt gtgtgtgtgt  5460
gtgtgtgaga gagagagaga gagagagaga aataggtcag atcagtatgt tgcaagcaga  5520
tctttagagc ctgtcaagtt ttggttagtt gtagtttcca tttggaaatg tagatgaatg  5580
cttgtagata ttggagattg tttctatctt gtaagaactt ttcactggtg ctgtaagcat  5640
ttcaaatagc accagtctta acctttaaat gggaagtaga aaaggtgagc ccaaagttta  5700
cagatgattt taatgctata catgttagtg tagtgatact tagaatgctt tgtttgatgt  5760
ttatttcaga aatgcgtata ctagaaaatc attttaatat taaaactggt gacttaaatac  5820
tagttgtaaa gtgtttttctt aaagaaggat cttggtactt aattgataaa gtgggtttag  5880
ataggagtag caagtgctct cgatagagaa agttttttgtt cacttcaata ctttgtcatt  5940
acaaccagtt cttcctgaaa atagttacat gtctagtaaa ttgatgtgaa attaactcgc  6000
tggagataat ctcattttag ttctgcaaat tctgcctggc tttaaaatgc atttttcataa  6060
tacttagaaa taatttgact gaaaataact gcttttttat ttaatcagtc aatcaactttt  6120
tactacaaat ttaattgagg gatttttttaa tttaattggt gctttaaaga agcaaataaa  6180
tccctgggtt ttgtttttctt cagtaaaatat cctaaagaaa ctctttaatg tatttgcgag  6240
tatatatata ttttcttatg catgctcgat gcattttcgt cctgagaaaa gtgttctcta  6300
cagaaactac ccgtgtgtta aaagaagatt ggcttaaaat ggctactgtg atgggaacag  6360
tgtcttaggg agatgcagct tggacttgag gtaaattgaa tactttacaa actggcttag  6420
agttttgctt taatgtcatt atatgtaaaa gggcacatga ttattgtaat tttgtattct  6480
ttatggtttc cttaattaaa ataataaatg tacagtgatt act                     6523
```

```
SEQ ID NO: 20              moltype = DNA   length = 2552
FEATURE                    Location/Qualifiers
source                     1..2552
                           mol_type = genomic DNA
                           organism = Mus musculus
SEQUENCE: 20
gcaagaatgt gatagagcct cgacggccgc catcttcttc ctttcttagc agttaaccga   60
gagcggcgcc tccctccgcg aggaaacggc agcgtgtgct gtagcgtggg tatggccgac  120
tactccacag tgcctccgcc gtcgtctggc tcggctggcg gcggcggcgg cggagtcgtt  180
aacgacgctt tcaaagatgc cctgcagaga gcgcggcaga ttgcagcaaa aattgggggt  240
gatgctggta catcattgaa ttcaaatgac tatggttatg ggggacaaaa aagaccctta  300
gaagatggag atcagccaga tgctaagaaa gtacctcccc aaaatgactc ttttggagca  360
cagttacctc caatgcatca gcagcaaagc agatctgtaa tgacagaaga atacaaagtc  420
ccagatggga tggtcggatt tataattggc agaggaggtg aacagatctc acgaatacag  480
caggaatctg gatgcaagat acagatagca cctgatagtg gtggcctacc agaaaggtct  540
tgtatgctaa ctggaacacc tgaatctgtc caatcagcaa aaagattatt ggaccagatt  600
gttgaaaagg gaagaccagc ccctggcttt catcatggtg atggacctgg aaatgcagtt  660
caggaaatca tgattccagc cagcaaagca ggactagtta ttggaaaggg gggcgagact  720
attaaacaac ttcaggaacg ggctggtgtt aaaatgatca tgattcaaga tgggcctcaa  780
aacactggtg ctgataaacc tcttaggatt acgggtgacc catacaaagt tcagcaagcc  840
aaggaaatgg tattagagtt aattcgtgat caaggtggtt tcagagaagt gcggaatgag  900
tatggctcaa gaataggagg caatgaaggg atagatgtcc caattccaag atttgctgtt  960
ggcattgtaa taggaaagaaa attaaaaaaa tacaaaatga tgctggttgtt             1020
cgaattcagt ttaagccaga tgatggaaca acacctgata ggatagcaca gataacagga  1080
cctccagaca ggtgtcagca cgctgcagaa ataatcacag accttctacg aagtgttcag  1140
gctggcaatc ctggtggacc gggacctggt ggtcgaggac gaggtagagg tcaaggaaac  1200
tggaatatgg ggccccgggg tggactccag gagtttaatt tcattgtgcc aactgggaaa  1260
actggactga tcattggaaa aggaggtgaa accataaaaa gcataagcca acagtctggt  1320
gcaagaatag aactgcagag aagccctcca cctaatgcag atcccaatat gaagttattt  1380
acaattcggg gcactccaca gcaaatagac tatgctcgac aactcataga agagaagatt  1440
gggggggccag taaatccttt agggccacct gtaccccatg ggcccatgg ggttccaggt  1500
cctcatgggc ctcctggacc tccagggccg ggaactccaa tgggaccata caaccctgg  1560
ccttacaatc caggaccacc tggcccagct cctcacggtc ctccagcccc atatgctccc  1620
cagggatggg gaaatgcgta tccacattgg cagcaacagg ctcctcctga cccagccaag  1680
gcaggagcag atccaaactc ggcagcttgg gctgcttatt atgctcacta ttaccagcag  1740
caggcacaac cccacctgc agctcctgcc ggtgcaccag ctacaacccca aacgaacgga  1800
caaggtaact acgagatca gcaggctcca gctccagctg acaggttga ttatacaaag  1860
gcttggaag aatactacaa gaaaatgggg caagcagttc ctgctcctgc tggggcccca  1920
ccaggtggtc agccggatta tagtgcagcc tgggctgagt actatagaca gcaagcagcg  1980
tattatgccc agacaagtcc ccaggggatg ccgcagcatc ctccagcacc tcagggattt  2040
gcaaatcatg caagaagcca ccatcattta tattaaccag tttttctttc ttaaaggatt  2100
cactcctgaa ttagctccat ttaaaggatt ttctttaact ttttgtgtat ctcttatgta  2160
tctcttctgc acagggccaa taataagaag tggacaatac agtatttgct tcattgtgtg  2220
```

-continued

```
gggggaaaaaa cctttgttaa atatatggat gcagacgact tgatgaagat cttaattttg   2280
ttttttggttt aaagtagtgt tttttttcccc cccttttttt ttattttgaa aatgtacaaa   2340
ataactatca ctactgatag gaggttaata tttctgtgta gaaatgaaaa ttggtttgtt   2400
tttagtctttt agtgtagatg tacacattcc agcaaatgta tttgcaacta ttatgtggtc   2460
gatgctttgt gatataaatg tacttttttca atgtatactt tcactttttaa aatgcctgtt   2520
ttgtgctttta caataaatga tatgaaacct cc                                  2552


SEQ ID NO: 21          moltype = RNA   length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = genomic RNA
                       organism = Homo sapiens
SEQUENCE: 21
gacaaacctc ttaggatta                                                    19


SEQ ID NO: 22          moltype = RNA   length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = genomic RNA
                       organism = Homo sapiens
SEQUENCE: 22
gagaagttcg gaatgagta                                                    19


SEQ ID NO: 23          moltype = RNA   length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = genomic RNA
                       organism = Homo sapiens
SEQUENCE: 23
gaaaggatag cacaaataa                                                    19


SEQ ID NO: 24          moltype = RNA   length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = genomic RNA
                       organism = Homo sapiens
SEQUENCE: 24
aataagaagt ggacaatac                                                    19


SEQ ID NO: 25          moltype = DNA   length = 21
FEATURE                Location/Qualifiers
misc_feature           1..21
                       note = primer
source                 1..21
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 25
tgtccagtca gcaaaacggt t                                                 21


SEQ ID NO: 26          moltype = DNA   length = 21
FEATURE                Location/Qualifiers
misc_feature           1..21
                       note = primer
source                 1..21
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 26
gaactgcatt tcccggtcca t                                                 21


SEQ ID NO: 27          moltype = DNA   length = 21
FEATURE                Location/Qualifiers
misc_feature           1..21
                       note = primer
source                 1..21
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 27
ggagtgtgga ttcgcactcc t                                                 21


SEQ ID NO: 28          moltype = DNA   length = 20
FEATURE                Location/Qualifiers
misc_feature           1..20
                       note = primer
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
```

-continued

```
SEQUENCE: 28
agattgagat cttctgcgac                                             20


SEQ ID NO: 29          moltype = DNA   length = 27
FEATURE                Location/Qualifiers
misc_feature           1..27
                       note = Probe
source                 1..27
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 29
aggcaggtcc cctagaagaa gaactcc                                     27


SEQ ID NO: 30          moltype = DNA   length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = primer
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 30
ccgtgtgcac ttcgcttca                                              19


SEQ ID NO: 31          moltype = DNA   length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = primer
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 31
gcacagcttg gaggcttga                                              19


SEQ ID NO: 32          moltype = DNA   length = 23
FEATURE                Location/Qualifiers
misc_feature           1..23
                       note = Probe
source                 1..23
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 32
catggagacc accgtgaacg ccc                                         23


SEQ ID NO: 33          moltype = DNA   length = 15
FEATURE                Location/Qualifiers
misc_feature           1..15
                       note = oligonucleotide motif
source                 1..15
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 33
ctcattccga acttc                                                  15


SEQ ID NO: 34          moltype = DNA   length = 18
FEATURE                Location/Qualifiers
misc_feature           1..18
                       note = Synthetic oligonucleotide
source                 1..18
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 34
ataaccatag tcatttga                                               18


SEQ ID NO: 35          moltype = DNA   length = 17
FEATURE                Location/Qualifiers
misc_feature           1..17
                       note = Synthetic oligonucleotide
source                 1..17
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 35
cccataacca tagtcat                                                17


SEQ ID NO: 36          moltype = DNA   length = 18
FEATURE                Location/Qualifiers
misc_feature           1..18
```

-continued

```
                            note = Synthetic oligonucleotide
source                      1..18
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 36
gagccatcta cacataaa                                                 18

SEQ ID NO: 37               moltype = DNA  length = 18
FEATURE                     Location/Qualifiers
misc_feature               1..18
                            note = Synthetic oligonucleotide
source                      1..18
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 37
caagagccat ctacacat                                                 18

SEQ ID NO: 38               moltype = DNA  length = 18
FEATURE                     Location/Qualifiers
misc_feature               1..18
                            note = Synthetic oligonucleotide
source                      1..18
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 38
tgtccattta agaatcca                                                 18

SEQ ID NO: 39               moltype = DNA  length = 18
FEATURE                     Location/Qualifiers
misc_feature               1..18
                            note = Synthetic oligonucleotide
source                      1..18
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 39
ttgtgtccat ttaagaat                                                 18

SEQ ID NO: 40               moltype = DNA  length = 16
FEATURE                     Location/Qualifiers
misc_feature               1..16
                            note = Synthetic oligonucleotide
source                      1..16
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 40
tacctttgct gctgat                                                   16

SEQ ID NO: 41               moltype = DNA  length = 17
FEATURE                     Location/Qualifiers
misc_feature               1..17
                            note = Synthetic oligonucleotide
source                      1..17
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 41
gactatacct ttgctgc                                                  17

SEQ ID NO: 42               moltype = DNA  length = 20
FEATURE                     Location/Qualifiers
misc_feature               1..20
                            note = Synthetic oligonucleotide
source                      1..20
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 42
actttgtatt cttctgtcat                                               20

SEQ ID NO: 43               moltype = DNA  length = 18
FEATURE                     Location/Qualifiers
misc_feature               1..18
                            note = Synthetic oligonucleotide
source                      1..18
                            mol_type = other DNA
                            organism = synthetic construct
SEQUENCE: 43
actttgtatt cttctgtc                                                 18
```

-continued

```
SEQ ID NO: 44          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = Synthetic oligonucleotide
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 44
gattcaggtg ttccagtta                                              19


SEQ ID NO: 45          moltype = DNA  length = 17
FEATURE                Location/Qualifiers
misc_feature           1..17
                       note = Synthetic oligonucleotide
source                 1..17
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 45
ttcaaactta ctggaca                                                17


SEQ ID NO: 46          moltype = DNA  length = 17
FEATURE                Location/Qualifiers
misc_feature           1..17
                       note = Synthetic oligonucleotide
source                 1..17
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 46
tacatctatc ccttcat                                                17


SEQ ID NO: 47          moltype = DNA  length = 16
FEATURE                Location/Qualifiers
misc_feature           1..16
                       note = Synthetic oligonucleotide
source                 1..16
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 47
ttacatctat cccttc                                                 16


SEQ ID NO: 48          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
misc_feature           1..20
                       note = Synthetic oligonucleotide
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 48
cttcctatta caatgccaac                                             20


SEQ ID NO: 49          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = Synthetic oligonucleotide
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 49
atttcttcct attacaatg                                              19


SEQ ID NO: 50          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = Synthetic oligonucleotide
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 50
ccatttcttc ctattacaa                                              19


SEQ ID NO: 51          moltype = DNA  length = 17
FEATURE                Location/Qualifiers
misc_feature           1..17
                       note = Synthetic oligonucleotide
source                 1..17
```

-continued

```
                              mol_type = other DNA
                              organism = synthetic construct
SEQUENCE: 51
gtttaaaata cattgcc                                                       17


SEQ ID NO: 52             moltype = DNA  length = 19
FEATURE                   Location/Qualifiers
misc_feature              1..19
                          note = Synthetic oligonucleotide
source                    1..19
                          mol_type = other DNA
                          organism = synthetic construct
SEQUENCE: 52
gagtttaaaa tacattgcc                                                     19


SEQ ID NO: 53             moltype = DNA  length = 18
FEATURE                   Location/Qualifiers
misc_feature              1..18
                          note = Synthetic oligonucleotide
source                    1..18
                          mol_type = other DNA
                          organism = synthetic construct
SEQUENCE: 53
gcttttttatg gtttcacc                                                     18


SEQ ID NO: 54             moltype = DNA  length = 20
FEATURE                   Location/Qualifiers
misc_feature              1..20
                          note = Synthetic oligonucleotide
source                    1..20
                          mol_type = other DNA
                          organism = synthetic construct
SEQUENCE: 54
atgctttta tggtttcacc                                                     20


SEQ ID NO: 55             moltype = DNA  length = 19
FEATURE                   Location/Qualifiers
misc_feature              1..19
                          note = Synthetic oligonucleotide
source                    1..19
                          mol_type = other DNA
                          organism = synthetic construct
SEQUENCE: 55
ataatcaacc tgtccagct                                                     19


SEQ ID NO: 56             moltype = DNA  length = 18
FEATURE                   Location/Qualifiers
misc_feature              1..18
                          note = Synthetic oligonucleotide
source                    1..18
                          mol_type = other DNA
                          organism = synthetic construct
SEQUENCE: 56
ataatcaacc tgtccagc                                                      18


SEQ ID NO: 57             moltype = DNA  length = 17
FEATURE                   Location/Qualifiers
misc_feature              1..17
                          note = Synthetic oligonucleotide
source                    1..17
                          mol_type = other DNA
                          organism = synthetic construct
SEQUENCE: 57
ataatcaacc tgtccag                                                       17


SEQ ID NO: 58             moltype = DNA  length = 18
FEATURE                   Location/Qualifiers
misc_feature              1..18
                          note = Synthetic oligonucleotide
source                    1..18
                          mol_type = other DNA
                          organism = synthetic construct
SEQUENCE: 58
tataatcaac ctgtccag                                                      18


SEQ ID NO: 59             moltype = DNA  length = 19
```

-continued

```
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = Synthetic oligonucleotide
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 59
gtataatcaa cctgtccag                                           19


SEQ ID NO: 60          moltype = DNA   length = 16
FEATURE                Location/Qualifiers
misc_feature           1..16
                       note = Synthetic oligonucleotide
source                 1..16
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 60
ataatcaacc tgtcca                                              16


SEQ ID NO: 61          moltype = DNA   length = 17
FEATURE                Location/Qualifiers
misc_feature           1..17
                       note = Synthetic oligonucleotide
source                 1..17
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 61
tataatcaac ctgtcca                                             17


SEQ ID NO: 62          moltype = DNA   length = 18
FEATURE                Location/Qualifiers
misc_feature           1..18
                       note = Synthetic oligonucleotide
source                 1..18
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 62
gtataatcaa cctgtcca                                            18


SEQ ID NO: 63          moltype = DNA   length = 17
FEATURE                Location/Qualifiers
misc_feature           1..17
                       note = Synthetic oligonucleotide
source                 1..17
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 63
gtataatcaa cctgtcc                                             17


SEQ ID NO: 64          moltype = DNA   length = 16
FEATURE                Location/Qualifiers
misc_feature           1..16
                       note = Synthetic oligonucleotide
source                 1..16
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 64
tataatcaac ctgtcc                                              16


SEQ ID NO: 65          moltype = DNA   length = 17
FEATURE                Location/Qualifiers
misc_feature           1..17
                       note = Synthetic oligonucleotide
source                 1..17
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 65
cccattttct tgtagta                                             17


SEQ ID NO: 66          moltype = DNA   length = 18
FEATURE                Location/Qualifiers
misc_feature           1..18
                       note = Synthetic oligonucleotide
source                 1..18
                       mol_type = other DNA
                       organism = synthetic construct
```

-continued

```
SEQUENCE: 66
acccattttc ttgtagta                                                    18

SEQ ID NO: 67          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = Synthetic oligonucleotide
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 67
tacccatttt cttgtagta                                                   19

SEQ ID NO: 68          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
misc_feature           1..20
                       note = Synthetic oligonucleotide
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 68
atacccattt tcttgtagta                                                  20

SEQ ID NO: 69          moltype = DNA  length = 18
FEATURE                Location/Qualifiers
misc_feature           1..18
                       note = Synthetic oligonucleotide
source                 1..18
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 69
atacccattt tcttgtag                                                    18

SEQ ID NO: 70          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
                       note = Synthetic oligonucleotide
source                 1..19
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 70
catacccatt ttcttgtag                                                   19

SEQ ID NO: 71          moltype = DNA  length = 18
FEATURE                Location/Qualifiers
misc_feature           1..18
                       note = Synthetic oligonucleotide
source                 1..18
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 71
catacccatt ttcttgta                                                    18

SEQ ID NO: 72          moltype = DNA  length = 18
FEATURE                Location/Qualifiers
misc_feature           1..18
                       note = Synthetic oligonucleotide
source                 1..18
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 72
tggagctaat tcaggagt                                                    18

SEQ ID NO: 73          moltype = DNA  length = 20
FEATURE                Location/Qualifiers
misc_feature           1..20
                       note = Synthetic oligonucleotide
source                 1..20
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 73
aaatggagct aattcaggag                                                  20

SEQ ID NO: 74          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
misc_feature           1..19
```

-continued

```
                             note = Synthetic oligonucleotide
source                       1..19
                             mol_type = other DNA
                             organism = synthetic construct
SEQUENCE: 74
ttgtccactt cttattatt                                                 19


SEQ ID NO: 75               moltype = DNA   length = 18
FEATURE                     Location/Qualifiers
misc_feature                1..18
                             note = Synthetic oligonucleotide
source                       1..18
                             mol_type = other DNA
                             organism = synthetic construct
SEQUENCE: 75
ccccacacaa tgaagcaa                                                  18


SEQ ID NO: 76               moltype = DNA   length = 19
FEATURE                     Location/Qualifiers
misc_feature                1..19
                             note = Synthetic oligonucleotide
source                       1..19
                             mol_type = other DNA
                             organism = synthetic construct
SEQUENCE: 76
ttcatcaagt cgtctgcat                                                 19


SEQ ID NO: 77               moltype = DNA   length = 18
FEATURE                     Location/Qualifiers
misc_feature                1..18
                             note = Synthetic oligonucleotide
source                       1..18
                             mol_type = other DNA
                             organism = synthetic construct
SEQUENCE: 77
gatcttcatc aagtcgtc                                                  18


SEQ ID NO: 78               moltype = DNA   length = 19
FEATURE                     Location/Qualifiers
misc_feature                1..19
                             note = Synthetic oligonucleotide
source                       1..19
                             mol_type = other DNA
                             organism = synthetic construct
SEQUENCE: 78
atattaacct cctatcagt                                                 19


SEQ ID NO: 79               moltype = DNA   length = 19
FEATURE                     Location/Qualifiers
misc_feature                1..19
                             note = Synthetic oligonucleotide
source                       1..19
                             mol_type = other DNA
                             organism = synthetic construct
SEQUENCE: 79
aatattaacc tcctatcag                                                 19


SEQ ID NO: 80               moltype = DNA   length = 19
FEATURE                     Location/Qualifiers
misc_feature                1..19
                             note = Synthetic oligonucleotide
source                       1..19
                             mol_type = other DNA
                             organism = synthetic construct
```

-continued

```
SEQUENCE: 80
atttatatca caaagcatc                                              19

SEQ ID NO: 81            moltype = DNA   length = 18
FEATURE                  Location/Qualifiers
misc_feature             1..18
                         note = Synthetic oligonucleotide
source                   1..18
                         mol_type = other DNA
                         organism = synthetic construct
SEQUENCE: 81
aagtacattt atatcaca                                               18

SEQ ID NO: 82            moltype = DNA   length = 20
FEATURE                  Location/Qualifiers
misc_feature             1..20
                         note = Synthetic oligonucleotide
source                   1..20
                         mol_type = other DNA
                         organism = synthetic construct
SEQUENCE: 82
catttattgt aaagcacaaa                                             20

SEQ ID NO: 83            moltype = DNA   length = 18
FEATURE                  Location/Qualifiers
misc_feature             1..18
                         note = Synthetic oligonucleotide
source                   1..18
                         mol_type = other DNA
                         organism = synthetic construct
SEQUENCE: 83
atcatttatt gtaaagca                                               18

SEQ ID NO: 84            moltype = RNA   length = 18
FEATURE                  Location/Qualifiers
misc_feature             1..18
                         note = siRNA sense
source                   1..18
                         mol_type = other RNA
                         organism = synthetic construct
SEQUENCE: 84
gcacttcgct tcacctct                                               18

SEQ ID NO: 85            moltype = DNA   length = 21
FEATURE                  Location/Qualifiers
misc_feature             1..21
                         note = Primer
source                   1..21
                         mol_type = other DNA
                         organism = synthetic construct
SEQUENCE: 85
cgtctgtgcc ttctcatctg c                                          21

SEQ ID NO: 86            moltype = DNA   length = 20
FEATURE                  Location/Qualifiers
misc_feature             1..20
                         note = primer
source                   1..20
                         mol_type = other DNA
                         organism = synthetic construct
SEQUENCE: 86
gcacagcttg gaggcttgaa                                             20

SEQ ID NO: 87            moltype = DNA   length = 20
FEATURE                  Location/Qualifiers
misc_feature             1..20
                         note = primer
source                   1..20
                         mol_type = other DNA
                         organism = synthetic construct
SEQUENCE: 87
ccgtctgaac tatcctgccc                                             20

SEQ ID NO: 88            moltype = DNA   length = 20
FEATURE                  Location/Qualifiers
misc_feature             1..20
```

-continued

```
                      note = primer
source                1..20
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 88
gccgtagtcg gtgtactcgt                                    20


SEQ ID NO: 89        moltype = DNA  length = 43
FEATURE              Location/Qualifiers
misc_feature         1..43
                      note = primer
source                1..43
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 89
taatacgact cactataggg tttttcacct ctgcctaatc atc          43


SEQ ID NO: 90        moltype = DNA  length = 43
FEATURE              Location/Qualifiers
misc_feature         1..43
                      note = primer
source                1..43
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 90
taatacgact cactataggg tttttcacct ctgcctaatc atc          43
```

1. A FUBP1 inhibitor for use in the treatment and/or prevention of Hepatitis B virus (HBV) infection.

2. The FUBP1 inhibitor for the use of claim **1**, wherein the HBV infection is a chronic infection.
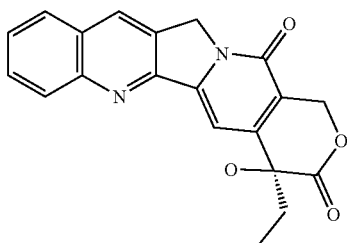
3. The FUBP1 inhibitor for the use of claim **1** or **2**, wherein the FUBP1 inhibitor is capable of reducing cccDNA and/or pgRNA in an infected cell.

4. The FUBP1 inhibitor for the use of any one of claims **1** to **3**, wherein the inhibitor prevents or reduces the FUBP1/FUSE interaction.
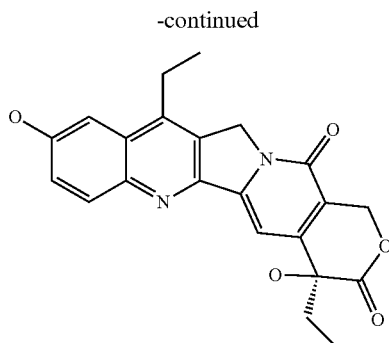
5. The FUBP1 inhibitor for the use of any one of claims **1** to **4**, wherein the inhibitor is selected from the compounds of Formula VII, IX or X
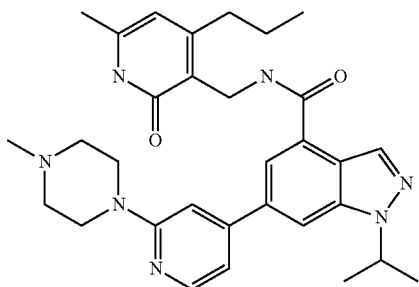
VIII

IX

-continued

X

6. The FUBP1 inhibitor for the use of any one of claims **1** to **3**, wherein said inhibitor is a nucleic acid molecule of 12 to 60 nucleotides in length comprising or consisting of a contiguous nucleotide sequence of 12 to 30 nucleotides in length which is at least 95% complementary to a mammalian FUBP1 target nucleic acid and capable of reducing FUBP1 mRNA.

7. The nucleic acid molecule for the use of claim **6**, wherein the nucleic acid molecule is selected from

a. a single stranded antisense oligonucleotide;

b. a siRNA molecule; or

c. a shRNA molecule.

8. The nucleic acid molecule for the use of claim **6** or **7**, wherein the mammalian FUBP1 target nucleic acid is selected from SEQ ID NO: 1 to 8

9. The nucleic acid molecule for the use of any one of claims **6** to **8**, wherein the contiguous nucleotide sequence is at least 98% complementarity to the target nucleic acid of SEQ ID NO: 1 and SEQ ID NO: 5.

10. The FUBP1 inhibitor for the use of any one of claims **1** to **9**, wherein the cccDNA in an HBV infected cell is reduced by at least 60% when compared to a control.

11. The nucleic acid molecule for the use of any one of claims **6** to **9**, wherein the FUBP1 mRNA is reduced by at least 60% when compared to a control.

12. A nucleic acid molecule of 12 to 60 nucleotides in length which comprises or consists of a contiguous nucleotide sequence of 12 to 30 nucleotides in length wherein the contiguous nucleotide sequence is at least 95% complementarity to a mammalian FUBP1 target nucleic acid.

13. The nucleic acid molecule of claim **12**, wherein the nucleic acid molecule is a siRNA or a single stranded antisense oligonucleotide.

14. The nucleic acid molecule of claim **12** or **13**, wherein the nucleic acid molecule comprises one or more 2' sugar modified nucleosides and one or more phosphorthioate linkages.

15. The nucleic acid molecule of any one of claims **12** to **14**, wherein contiguous nucleotide sequence is complementary to a target sequence selected from the group consisting of position 14200-14218, 14413-14431, 14966-14984 and 30344-30362 on SEQ ID NO: 1.

16. A conjugate compound comprising the nucleic acid molecule of any one of claims **12** to **15** and at least one conjugate moiety covalently attached to an oligonucleotide component of the nucleic acid molecule.

17. The conjugate compound of claim **16**, wherein the conjugate moiety is selected from one of the trivalent GalNAc moieties in FIG. **1**.

18. The conjugate compound of claim **16** or **17** comprising a physiologically labile linker composed of 2 to 5 linked nucleosides comprising at least two consecutive phosphodiester linkages, wherein the physiologically labile linker covalently bound at the 5' or 3' terminal of the oligonucleotide component.

19. A pharmaceutical composition comprising one or more nucleic acid molecule(s) of any one of claims **12** to **15**, or one or more of the conjugate compound(s) of any one of claims **16** to **18**, or acceptable salts thereof and a pharmaceutically acceptable diluent, carrier, salt and/or adjuvant.

20. An in vivo or in vitro method for modulating FUBP1 expression in a target cell which is expressing FUBP1, said method comprising administering the nucleic acid molecule of any one of claims **12** to **15**, or the conjugate compound of any one of claims **16** to **18** or the pharmaceutical composition of claim **19** in an effective amount to said cell.

21. The nucleic acid molecule of any one of claims **12** to **15**, or the conjugate compound of any one of claims **16** to **18** or the pharmaceutical composition of claim **19** for use as a medicament.

* * * * *