



US 20120174243A1

(19) United States

(12) Patent Application Publication
Combs et al.(10) Pub. No.: US 2012/0174243 A1
(43) Pub. Date: Jul. 5, 2012

- (54) SOGA POLYNUCLEOTIDES AND POLYPEPTIDES AND USES THEREOF
C12N 1/19 (2006.01)
C12N 15/63 (2006.01)
G01N 33/566 (2006.01)
A61K 38/17 (2006.01)
A61P 5/50 (2006.01)
A61P 3/10 (2006.01)
A61K 31/7088 (2006.01)
C07K 14/435 (2006.01)
C07K 16/18 (2006.01)
C07H 21/04 (2006.01)
A61K 39/395 (2006.01)
- (76) Inventors: Terry P. Combs, Chapel Hill, NC (US); James A. Swenberg, Pittsboro, NC (US)
- (21) Appl. No.: 13/376,239
- (22) PCT Filed: Jun. 4, 2010
- (86) PCT No.: PCT/US2010/037472
- (52) U.S. Cl. 800/13; 424/139.1; 435/6.1; 435/325; 435/348; 435/375; 435/419; 435/252.3; 435/252.33; 435/254.2; 435/320.1; 436/501; 514/1.1; 514/6.7; 514/6.8; 514/44 R; 514/44 A; 530/350; 530/387.9; 536/23.5; 536/24.5
- § 371 (c)(1),
(2), (4) Date: Feb. 15, 2012

Related U.S. Application Data

(60) Provisional application No. 61/184,392, filed on Jun. 5, 2009.

Publication Classification

- (51) Int. Cl.
A01K 67/027 (2006.01)
C12Q 1/68 (2006.01)
C12N 5/10 (2006.01)
C12N 5/02 (2006.01)
C12N 1/21 (2006.01)

(57)

ABSTRACT

The present invention relates to the identification of polynucleotides and polypeptides involved in insulin and adiponectin signaling and regulation of glucose production. The invention further relates to the use of the identified polynucleotides and polypeptides, and inhibitors of the polynucleotides and polypeptides, in the regulation of glucose production and the monitoring and treatment of metabolic disorders such as diabetes.

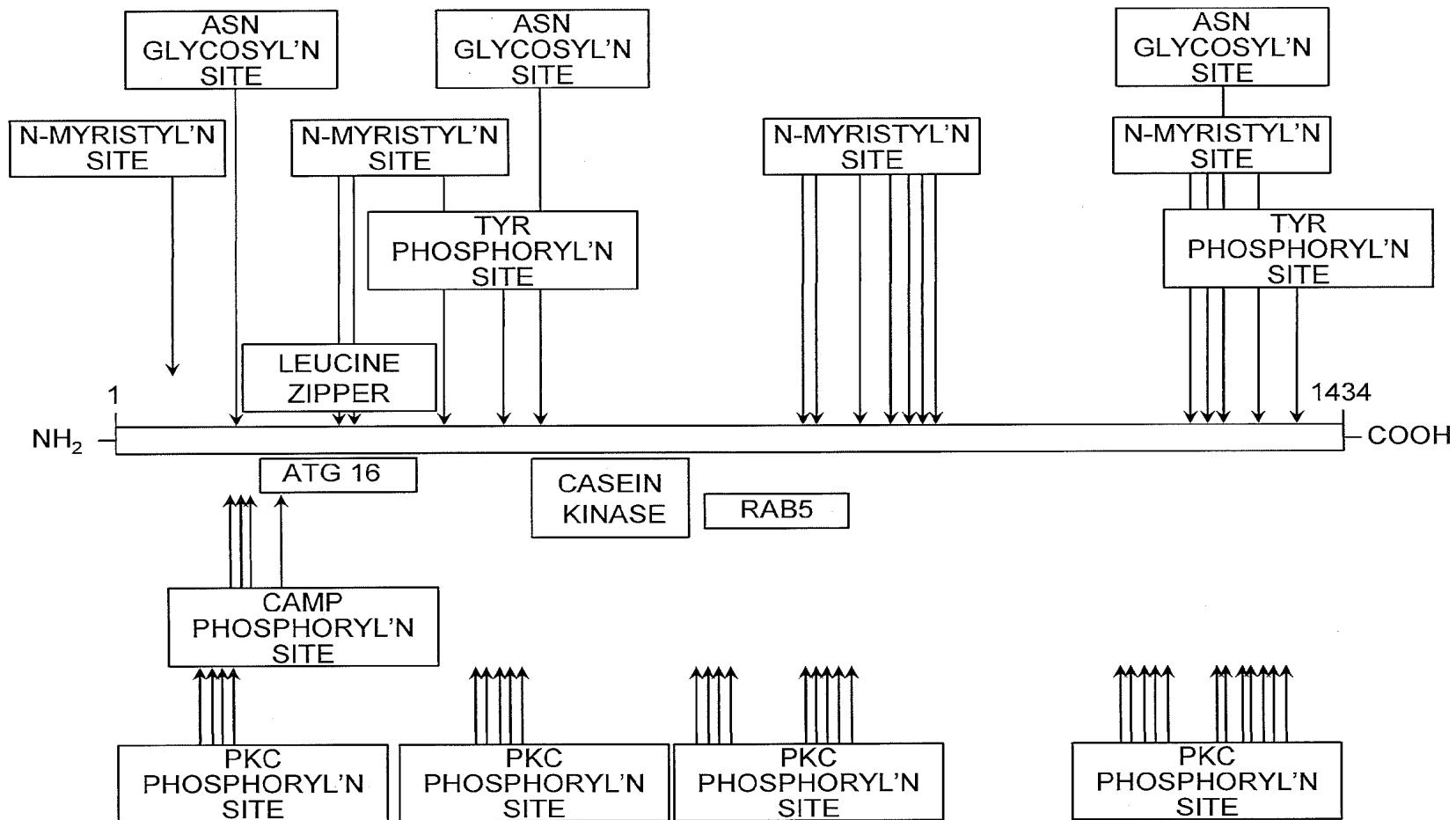


FIG. 1

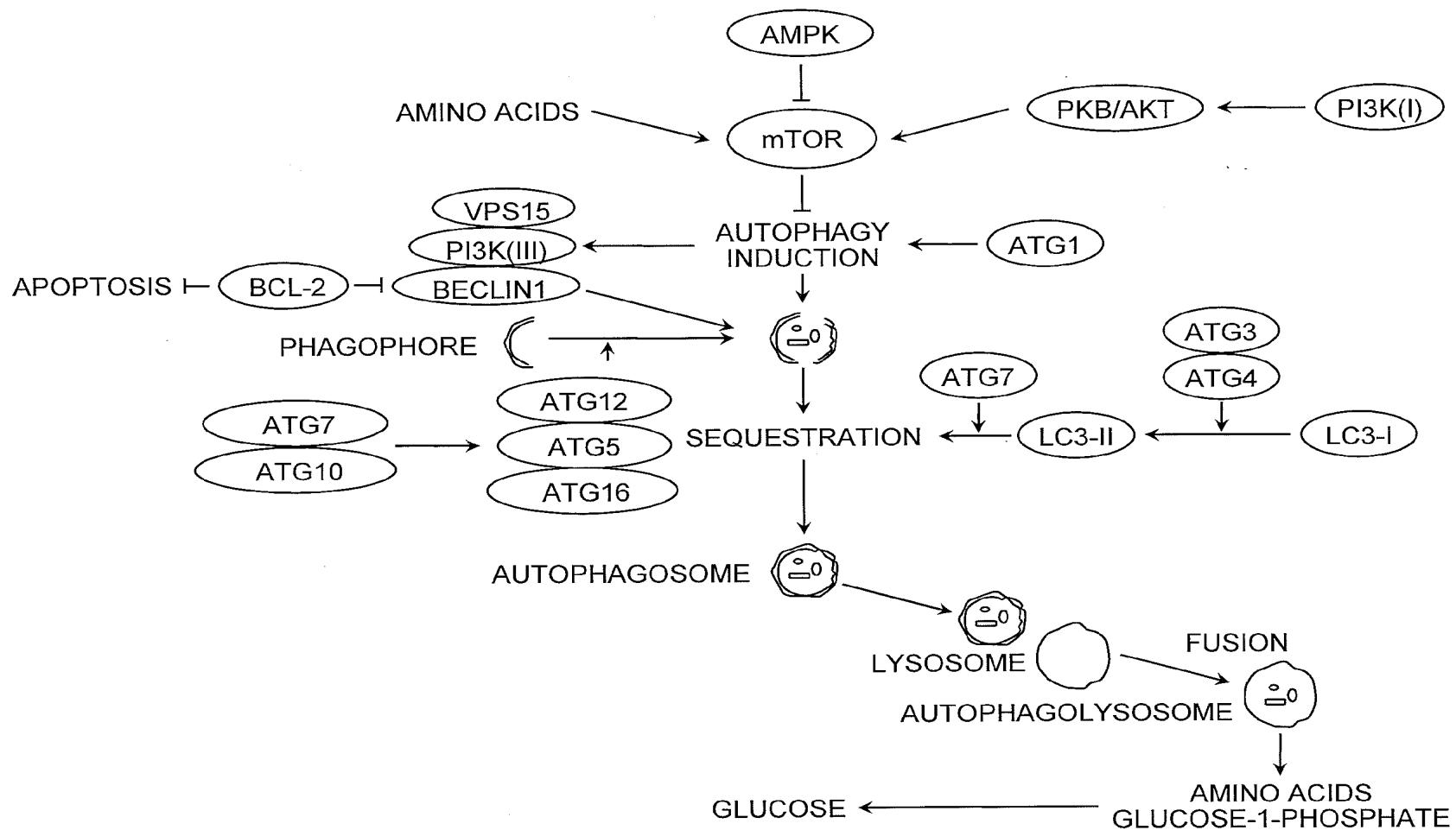


FIG. 2

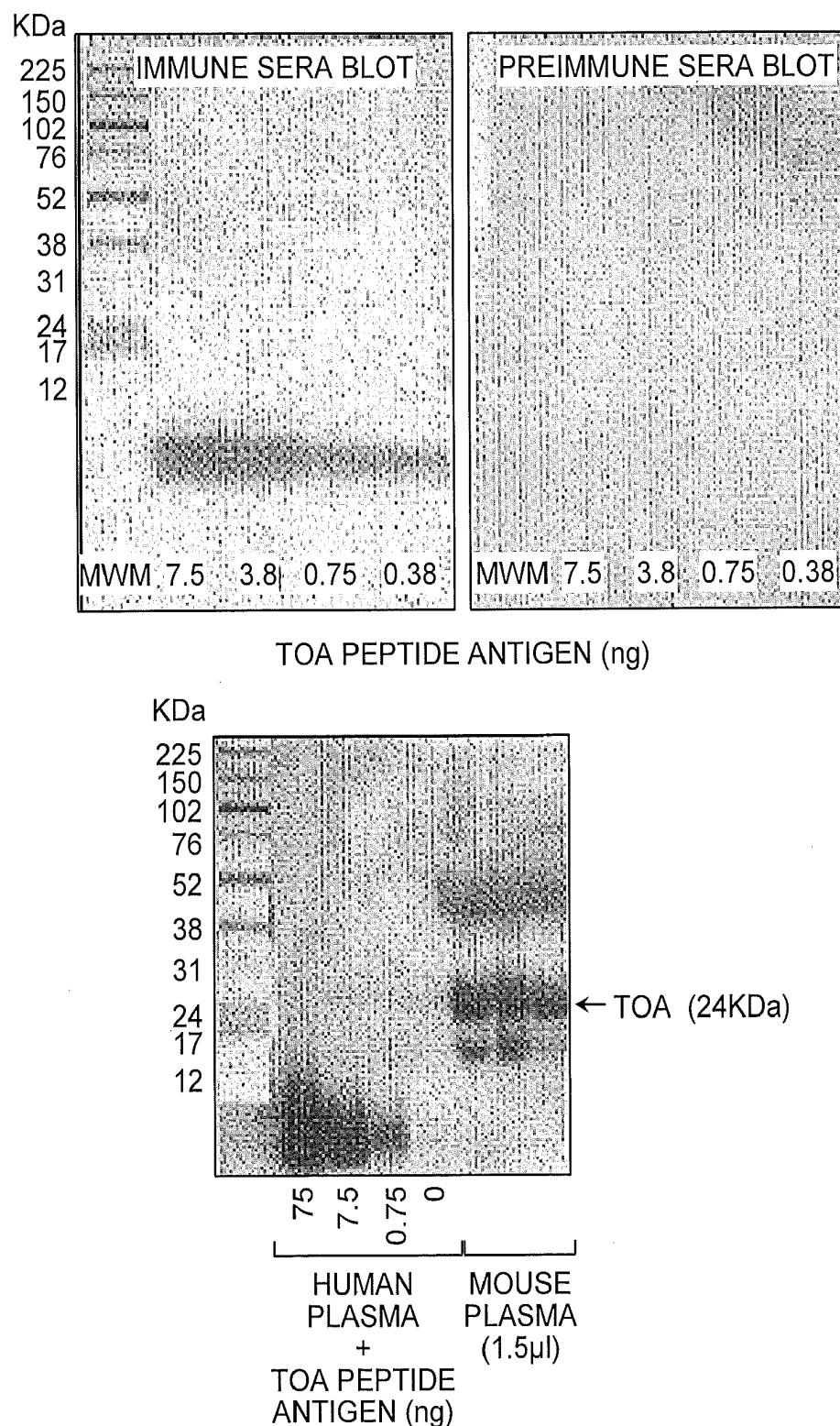


FIG. 3

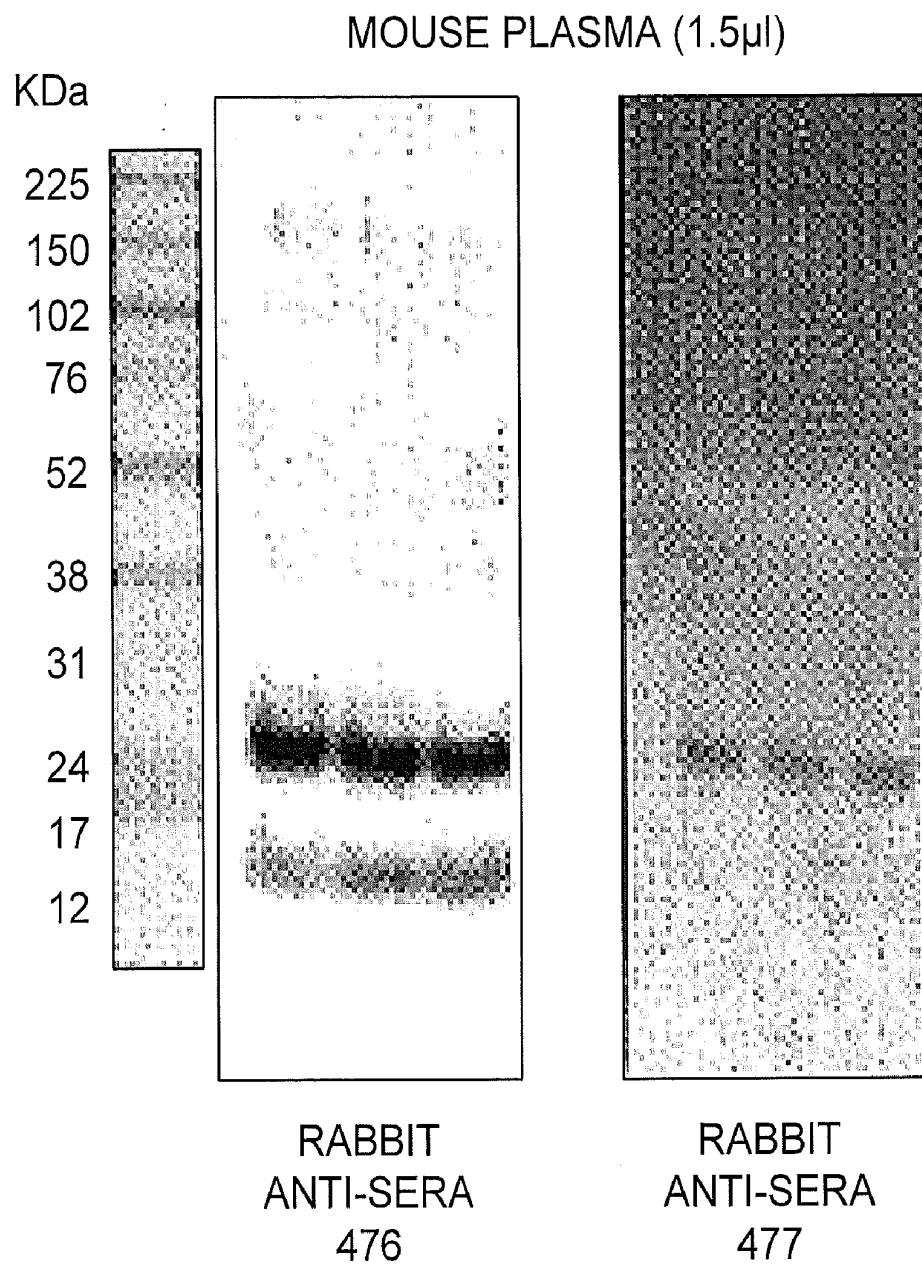


FIG. 4

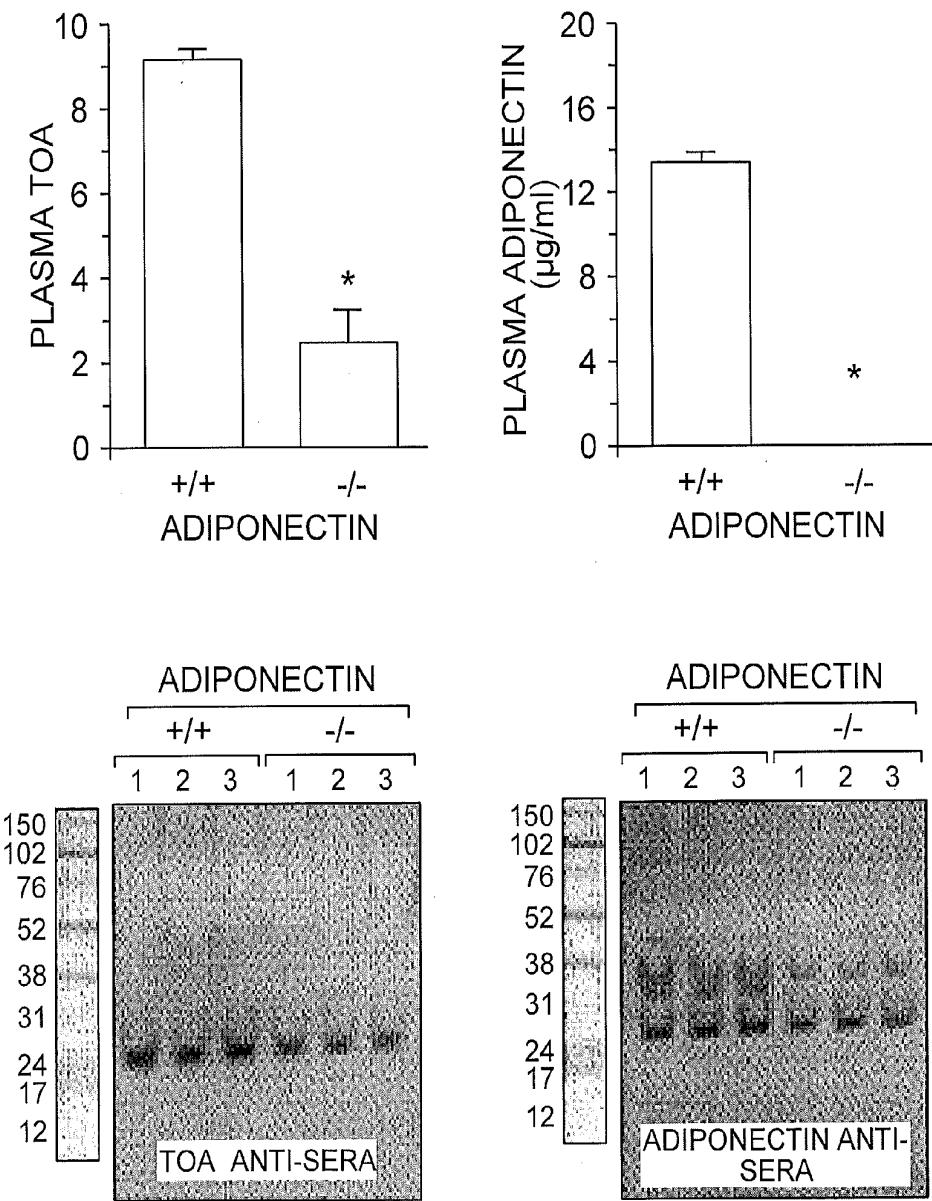


FIG. 5

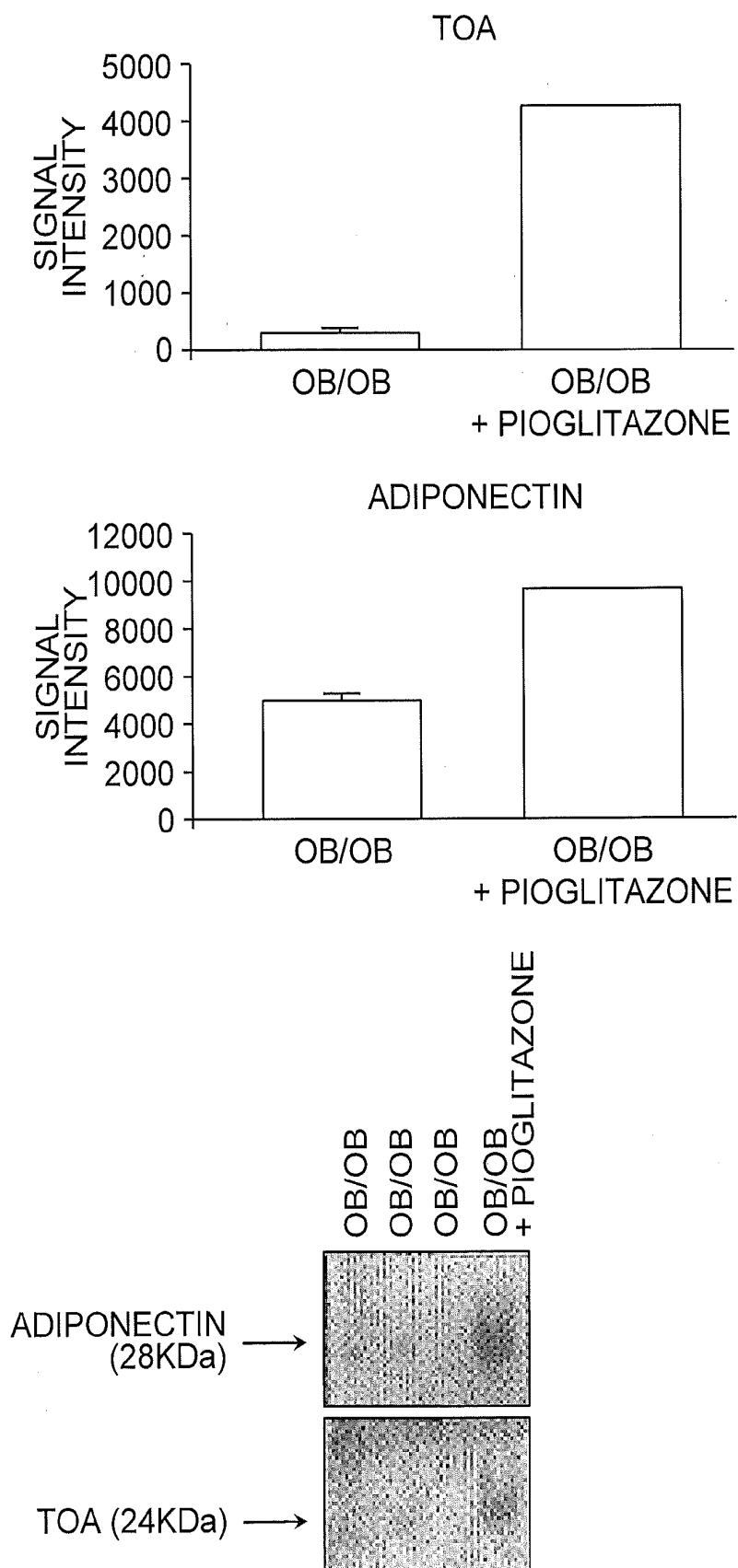


FIG. 6

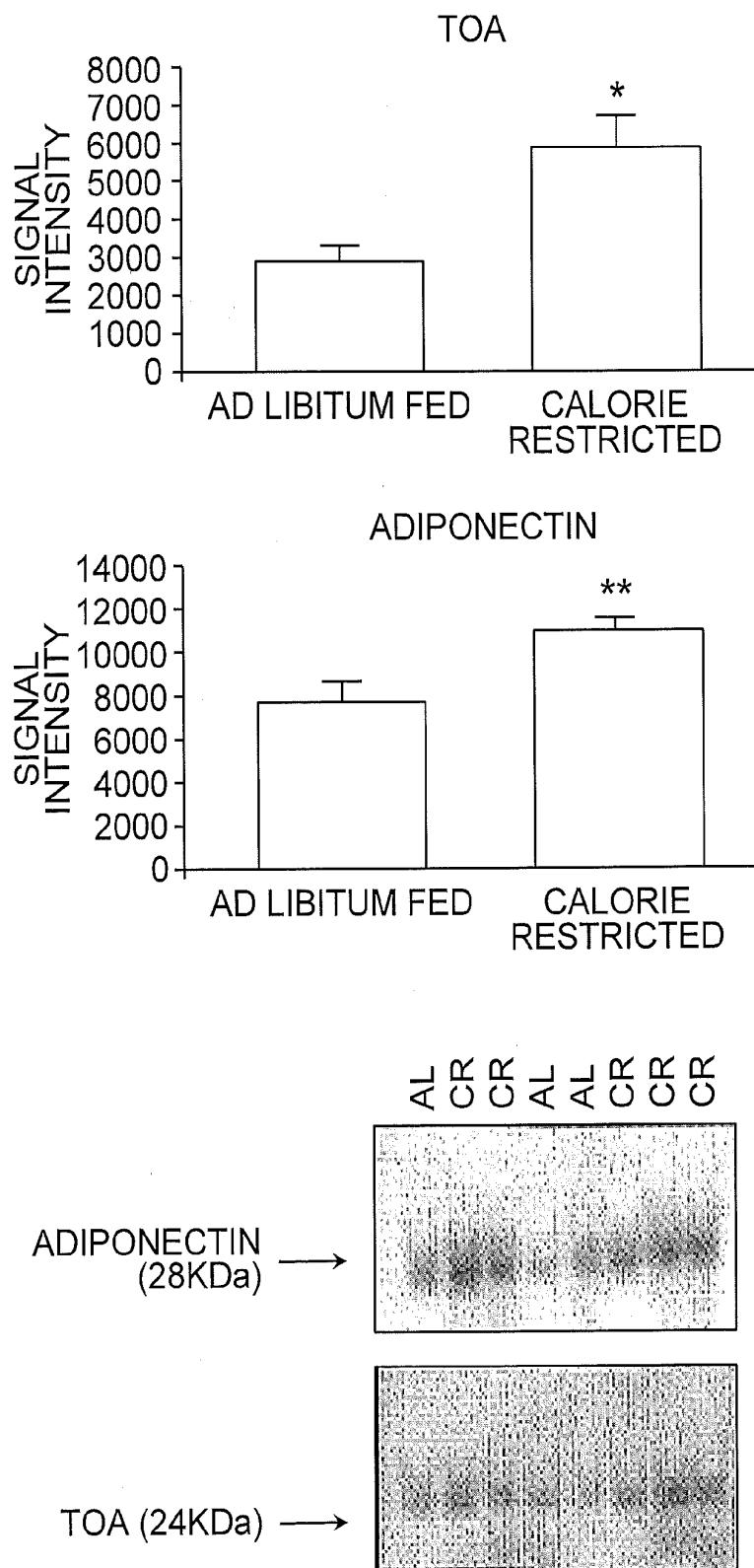


FIG. 7

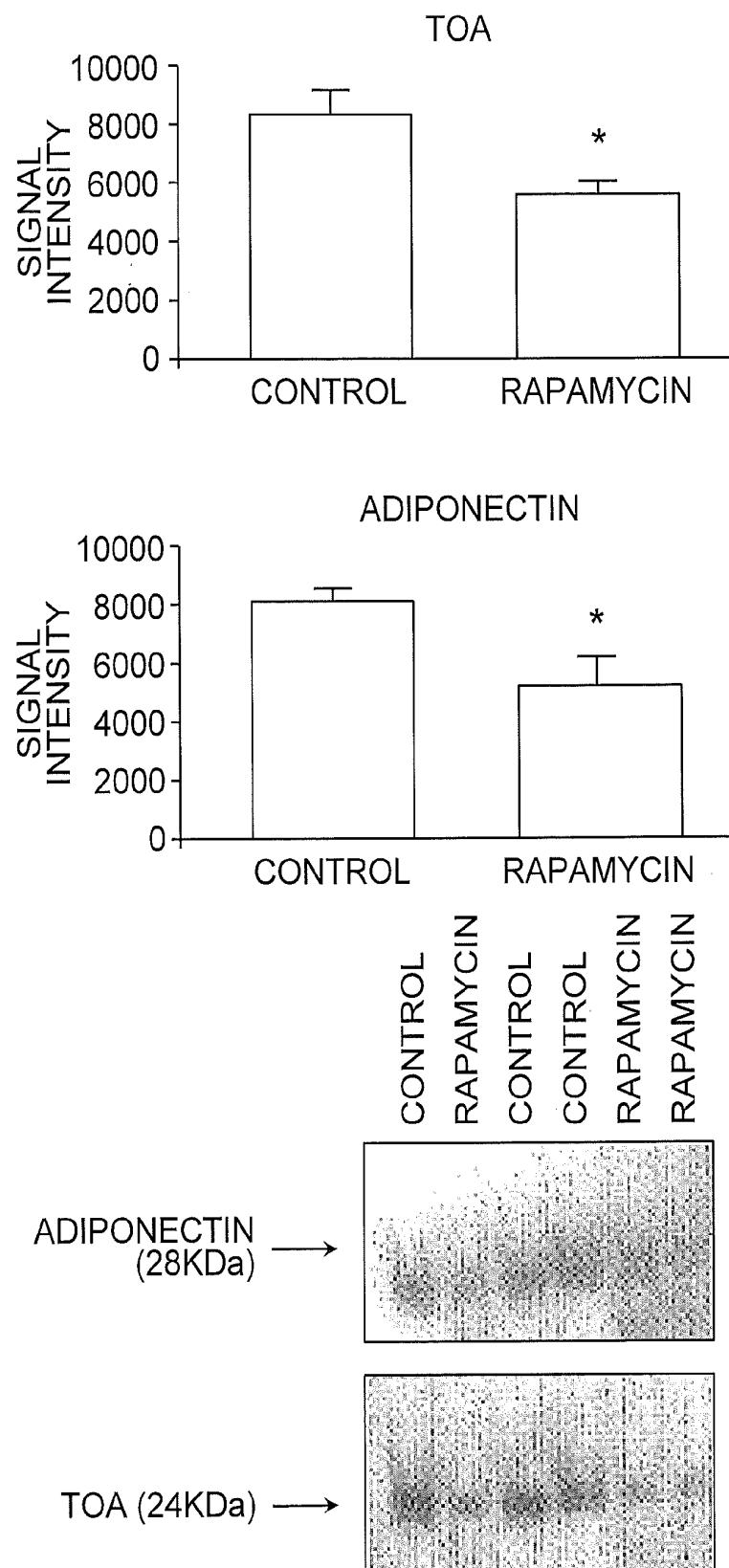


FIG. 8

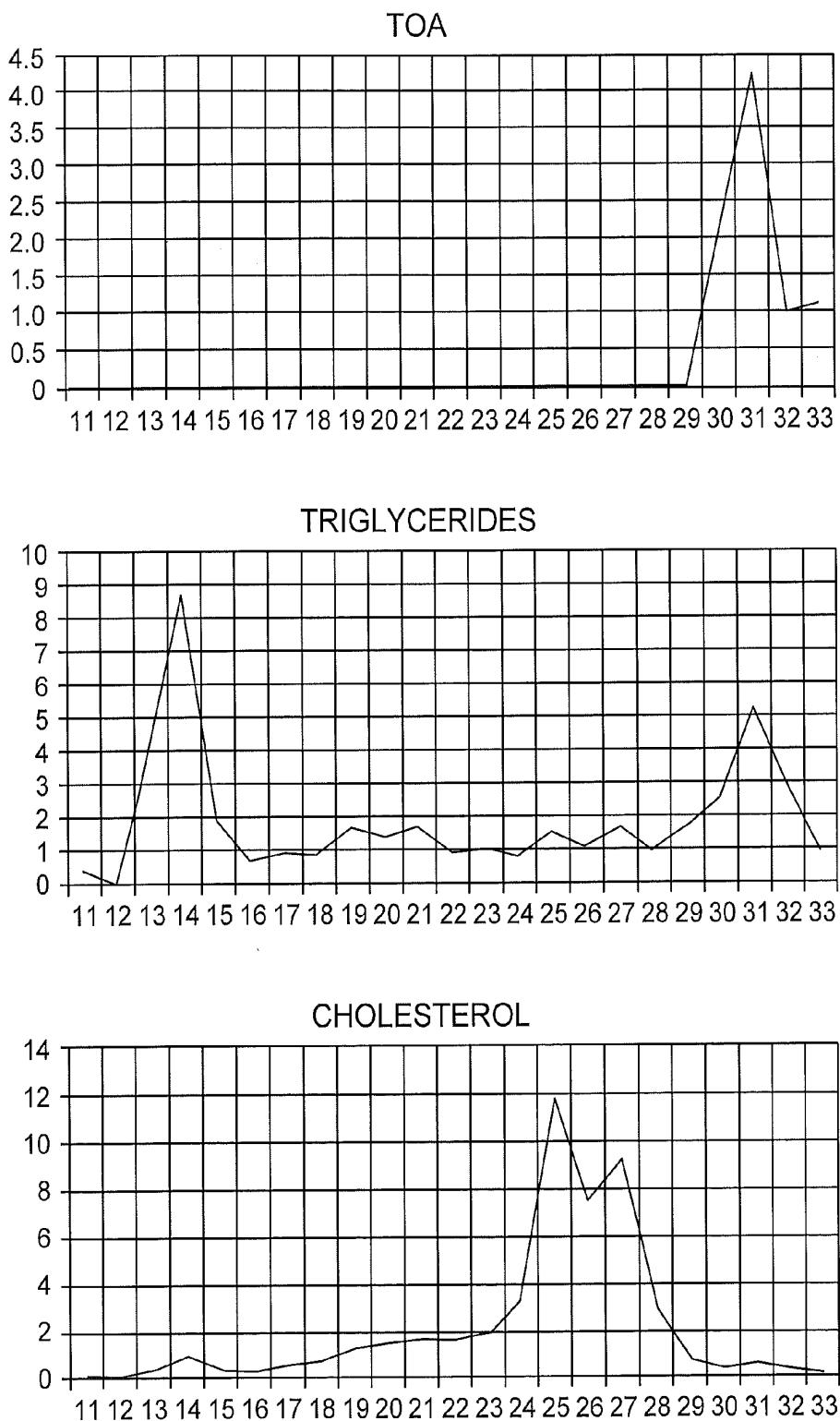
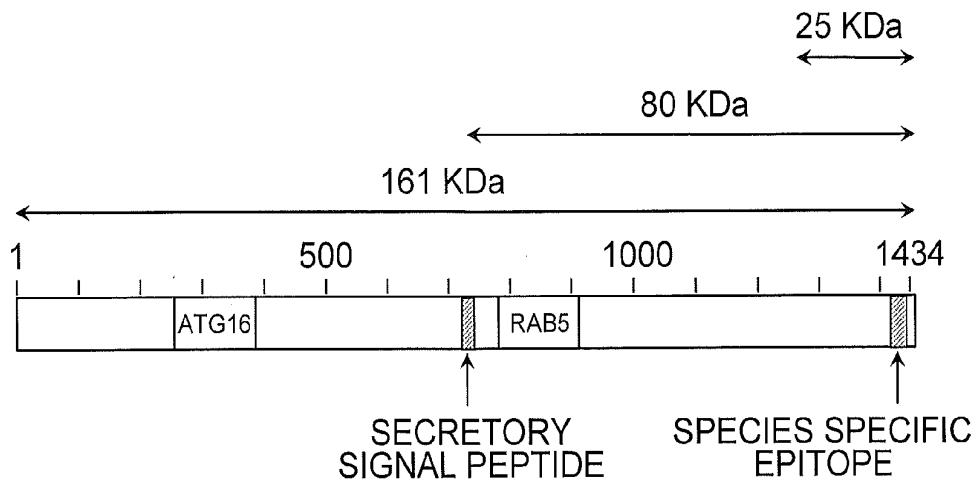


FIG. 9



1 MLDGPGGLVRELEELRSENDYLKDEIEELRAEMLEM RDVYMEEDVYQLQYRLR
 55 KAERRSLRAAQTGQVDGELIRGLEQDVVKVSKDISMRLHKELEVVEKKRMRLEEE
 109 NEGLRQRLIETELAKQLQTELDRPREHSLKKRGTRSLGKTDKKPTAQEDSAKL
 163 KCQLHFAKEEASALMCKKLTKLAKENDSMKEELLKYRSIYGDLDAALSAELADA
 217 PHSRETELKVHLKLVEEEANLLSRRIVELEVENRGLRAEMDDMKDHGGGGPEA
 271 RLAFSSLGECGESLAELRRHLQFVEEEAEELLRRSSAELDQNKLNNELAKYR
 325 SEHELDVTLSEDSCSVLSEPSQEELAAKLOQIGELSGKVKKLQYENRVLLSNLQ
 379 RCDSLASCQSTRPMLETDAEAGDSAQCVPAPLGETLEPHAARLCRAAEALPGL
 433 REQAALVSKAIDVLVADANGFSVGLRLCLDNECADRLHEAPDNSEGPRDAKLI
 487 HAILVRLSVLQQELNAFTRKADVALGSSGKEQPEPFALPALGSQGPAKEIMLS
 541 KDLGSDFQPPDFRDLLWEPRIAREAFRTGDLESKPDPSRNFRPYRAEDNDSYAS
 595 EIKDLQLVLAEAHDSLRLQEQQLSQRQLRKEEADSFNQKMVQLKEDQQRALLR
 649 REFELOQSLSLQRRLEQKFWSQEKNILVQESQQFKHNFLLFMKLRFWFLKWRQG
 703 **KVLPSEEDDFLEVNSMKELYLLMEEEEMNAQHSDNKACTGESWTQNTPNECIKT**
 757 **LADMKVTLKELCWLLQDERRGLTELQQQFAKAKATWETERAELKGHASQMLKA**
 811 **GKGASERPGPDWKALQREREQQHLLAESYSAVMELTRQLQLSERHWSQEKLQ**
 865 **LVERLQGEKQQVEQQVKELQNRLSQLQKAAEPWVLKHSDMEKQDNSWKEARSEK**
 919 THDKEGVSEAELGGTGLKRTKS VSSMSEFESLLDCSPYLAGGDARNKKLPNGPA
 973 FAFVSTEPVEPEKDAKEKAGLSTRDCSHIGSLACQEPAGRQMQRSYTAPDKTGI
 1027 RVYYSPPVARRLGVPVVHDKEGKILIEPGFLTTAKPKESAEDGLAESSYSRW
 1081 LCNFSRQRLDGGSGASTSGSGPAFPALHDFEMSGNMSDDMKEITNCVRQAMRSG
 1135 SLERKVKN TSSQTVGVATVGTQTIRT VSVGQTDPPRSSLHSKSWS PRSSLV S
 1189 VRSKQISSSLDKVHSRIERPCCSPKYGSPKLQRRSVSKLDSTKDRSLWNLHQGK
 1243 QNGSAWARSTT RDSPVLRNINDGLSSLFSVVEHGSTESVWKLGMSEARTKPE
 1297 PPKYGIVQEFFF RNVCGRAPSPTTAAGEESCKPEPLSPAS YHQPEGVSRILNKK
 1351 AAKAGGSEEVRPTMLSQVGKD GIL RDGD GSLIL PSED AVCDCSAQSLASCFIRP
 1405 SRNTIRHSPSKCRLHPSESGWGGEERAAPQ

FIG. 10

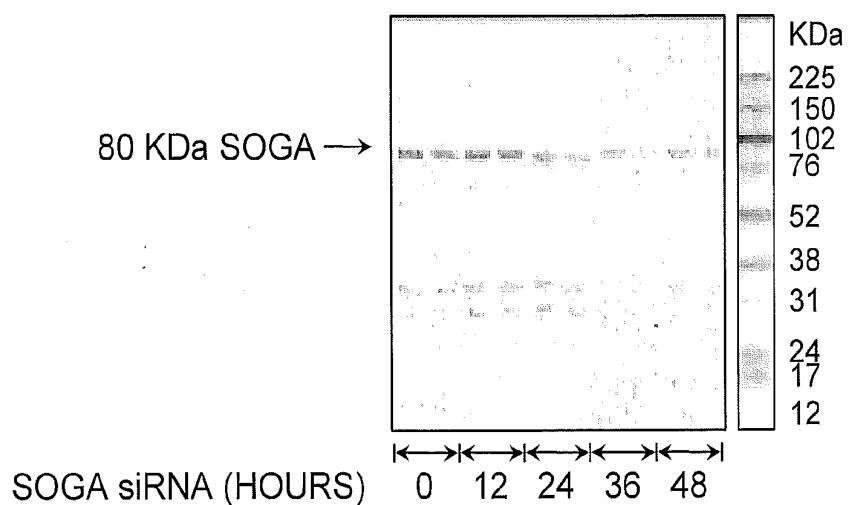


FIG. 11A

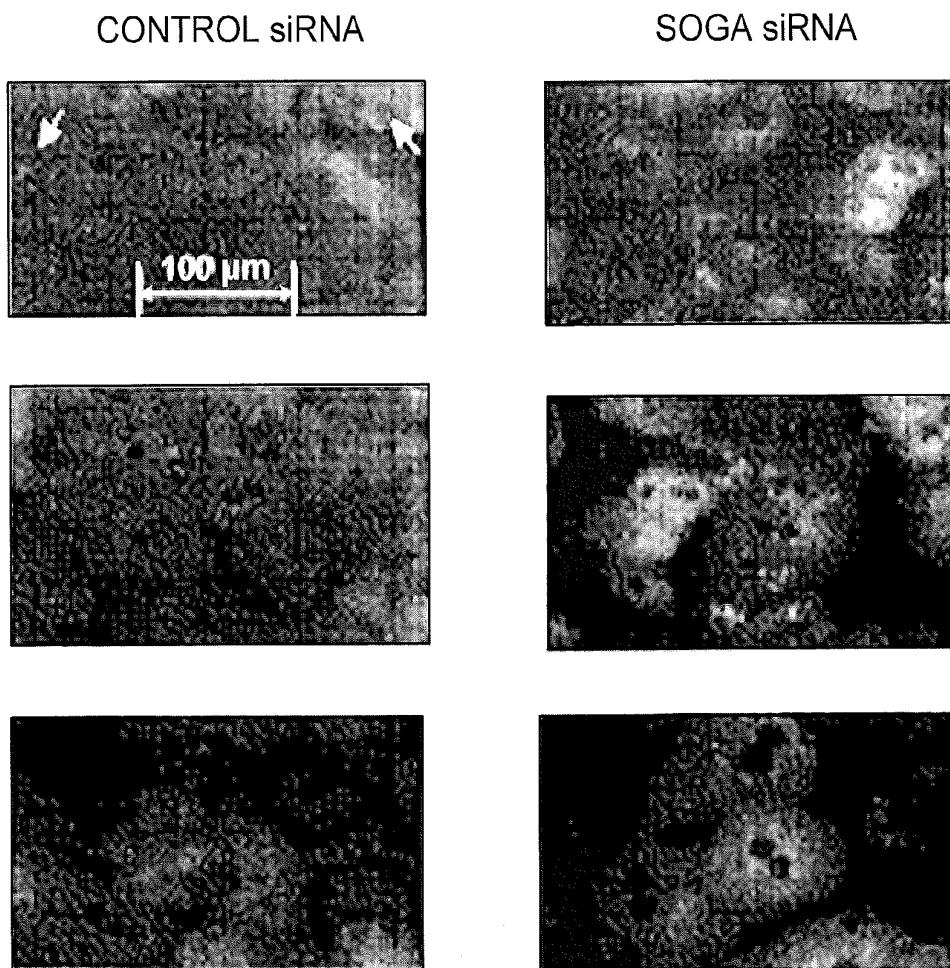


FIG. 11B

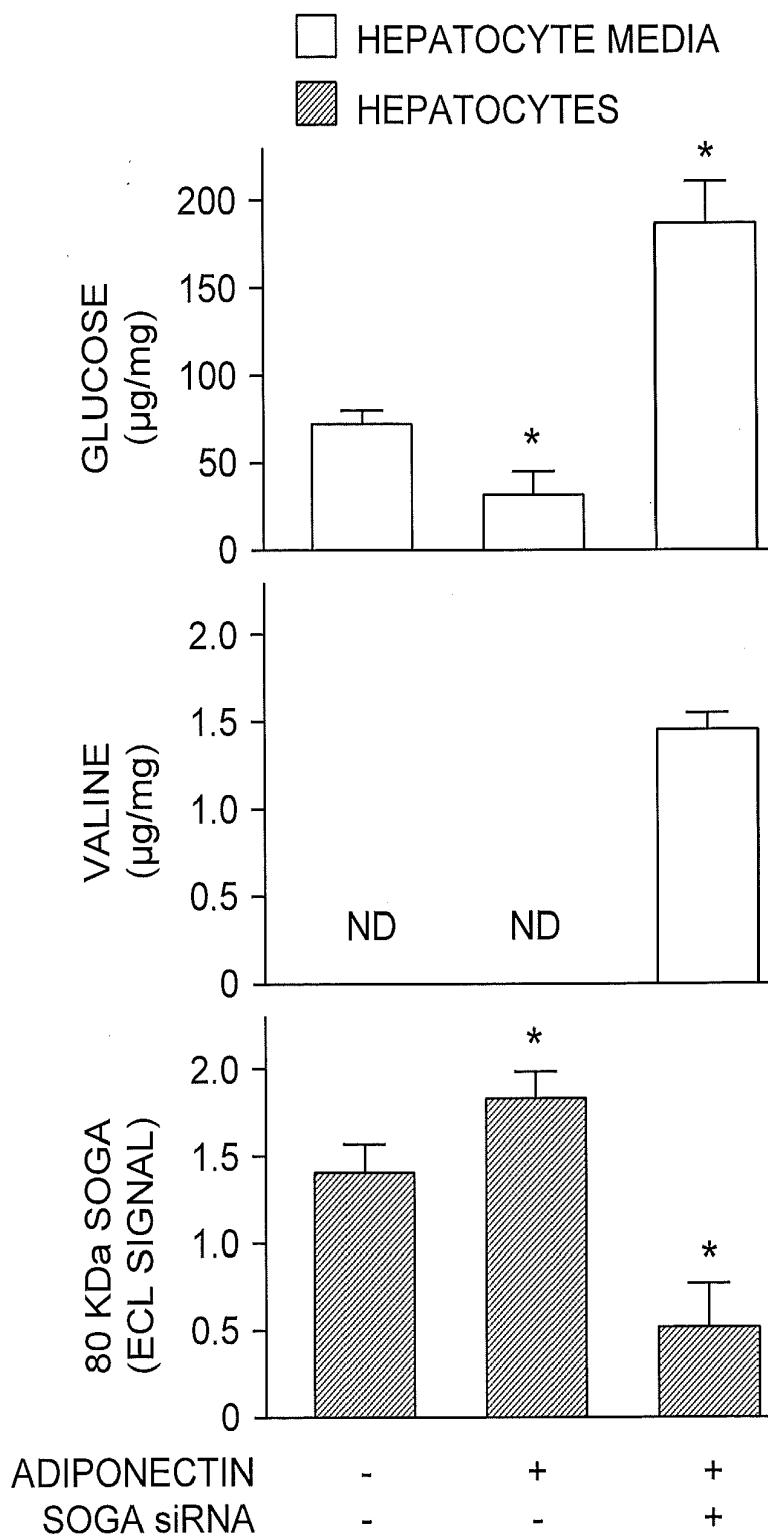


FIG. 11C

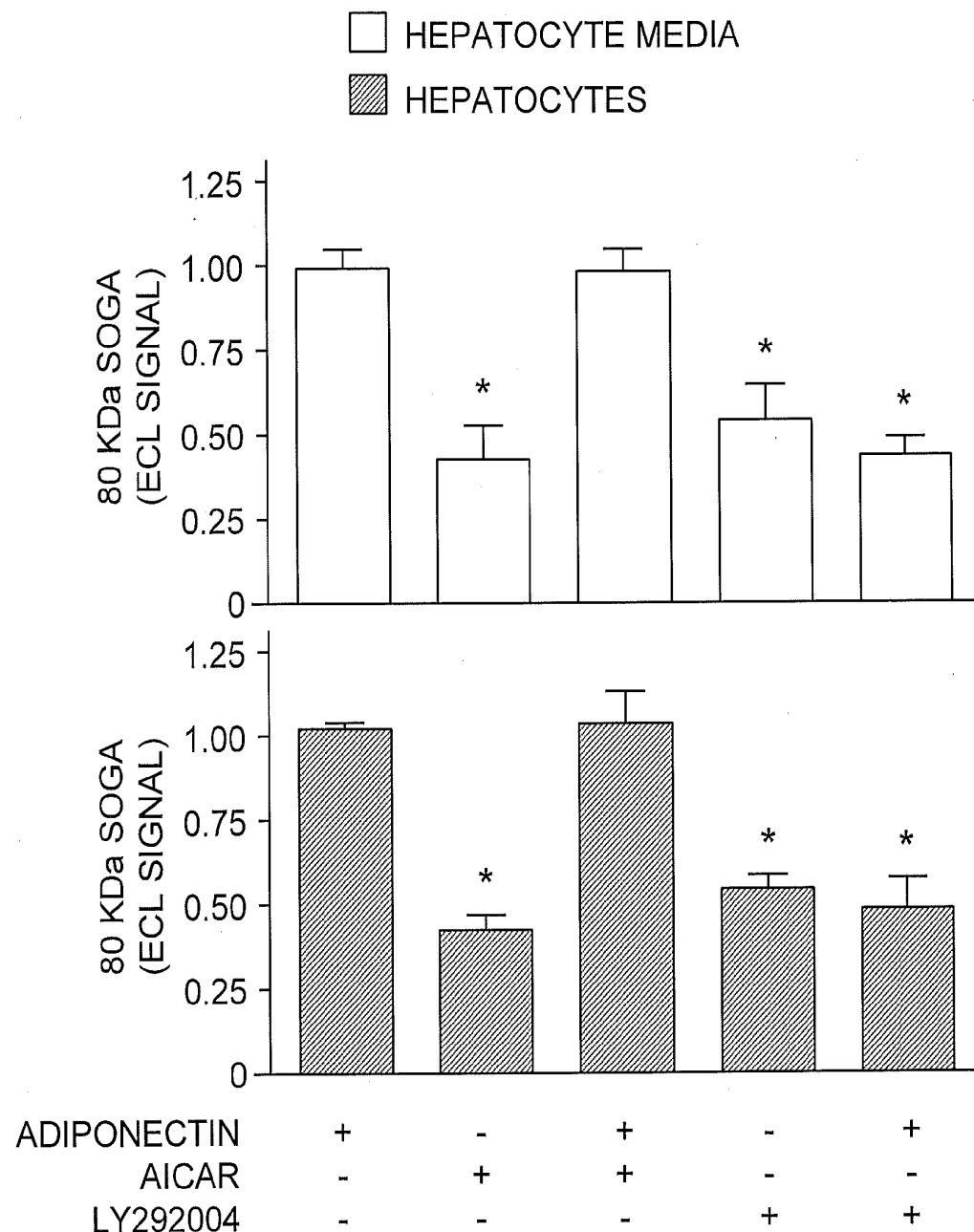


FIG. 11D

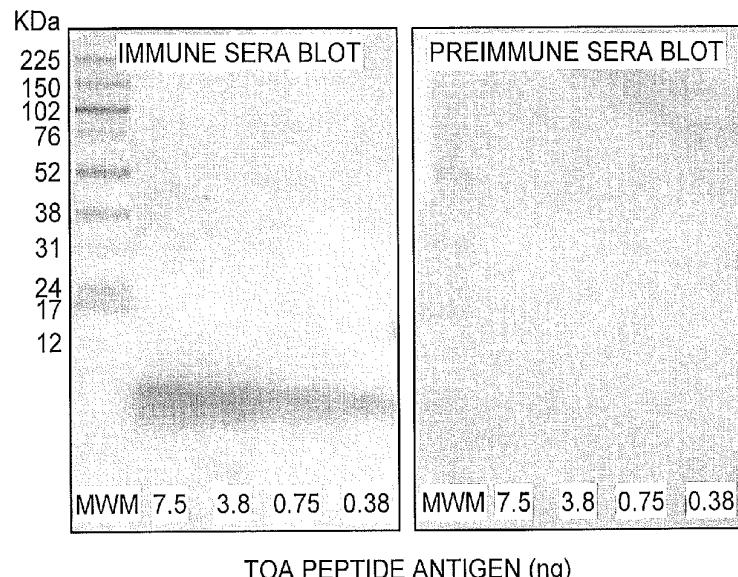


FIG. 12A

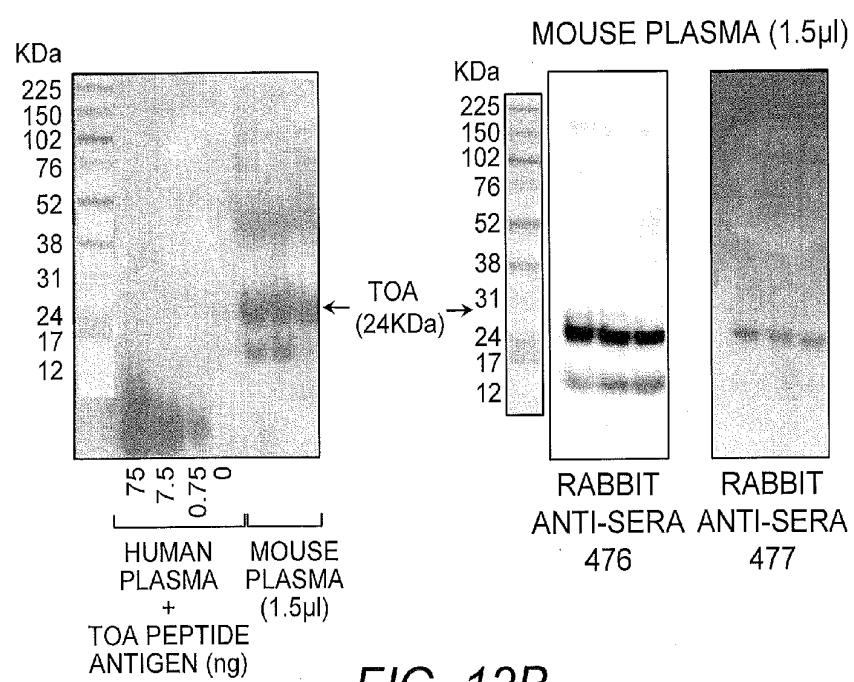


FIG. 12B

DAD1A, SIG=220.8 REF=360.40 (2009JUNE\PLASMA-06040906.D)

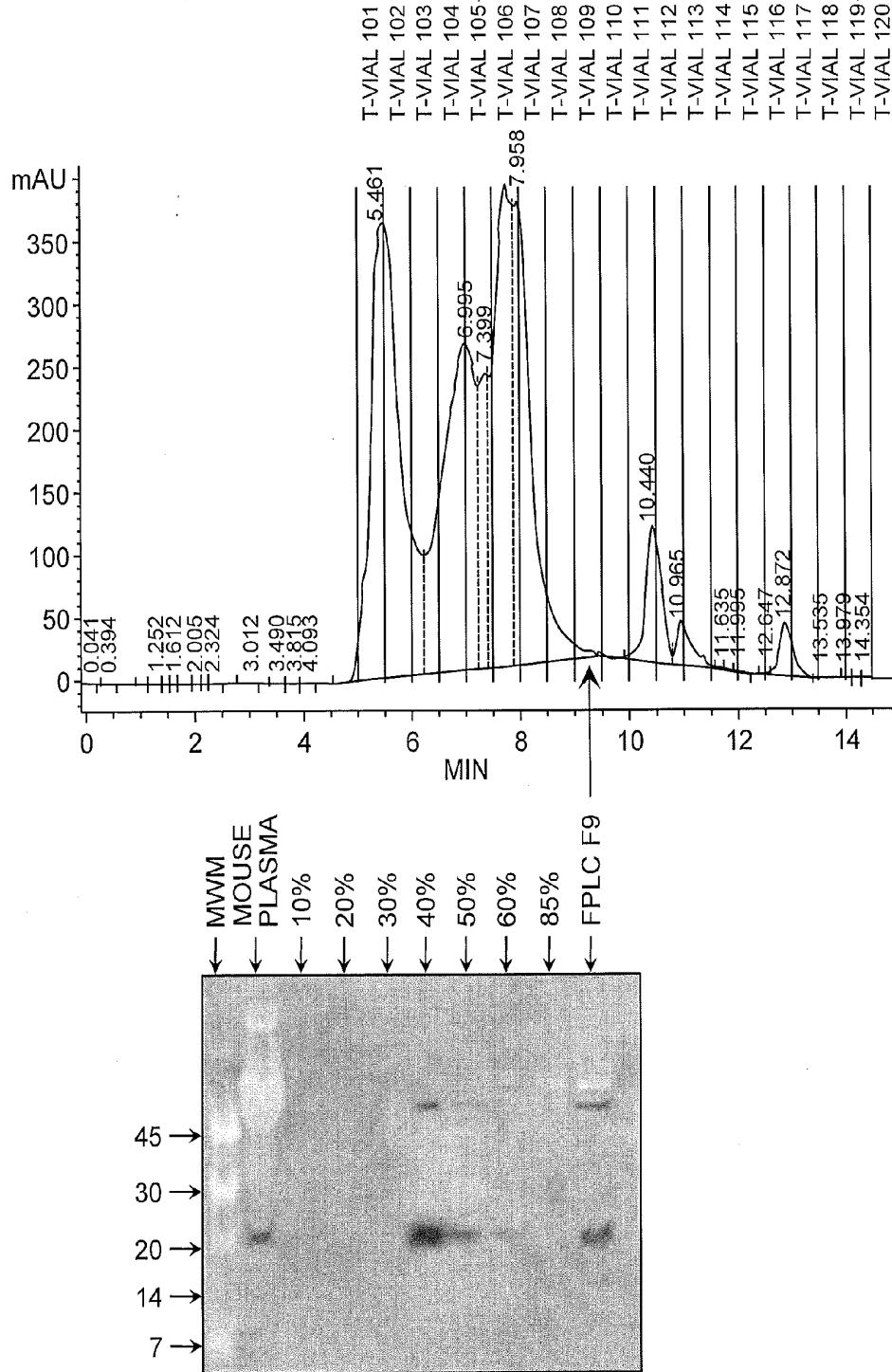


FIG. 12C

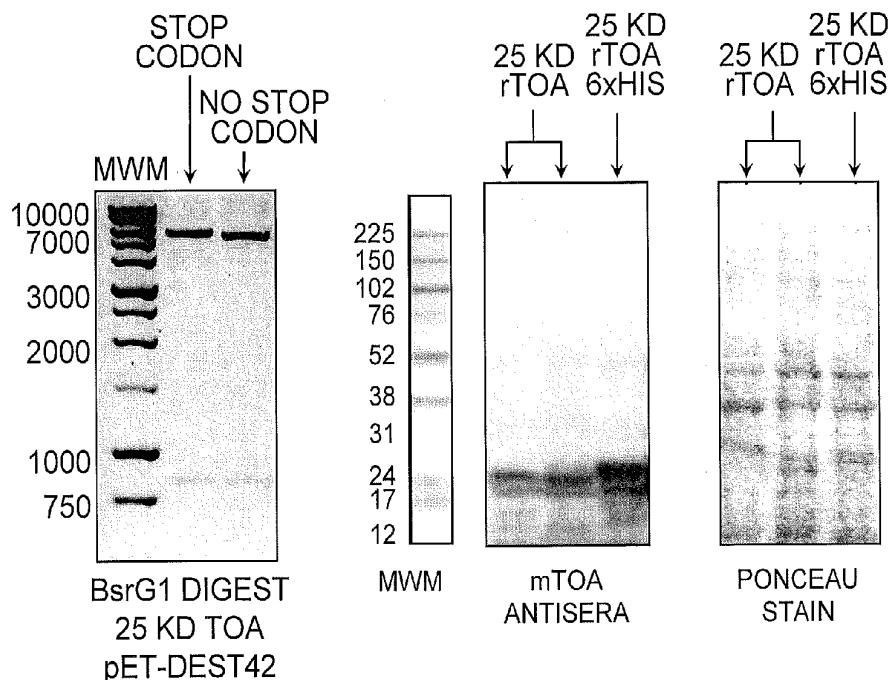


FIG. 13A

FIG. 13B

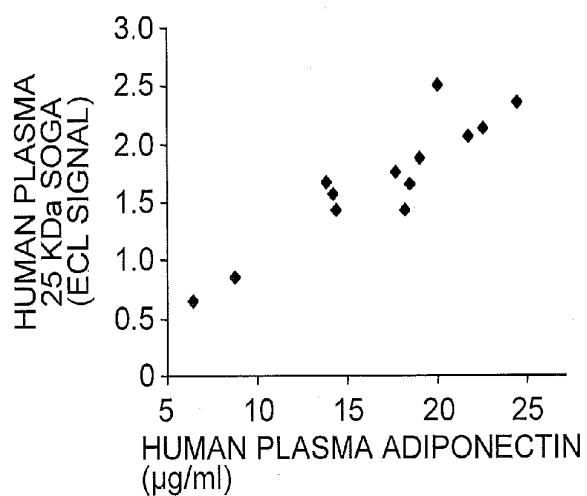


FIG. 14A

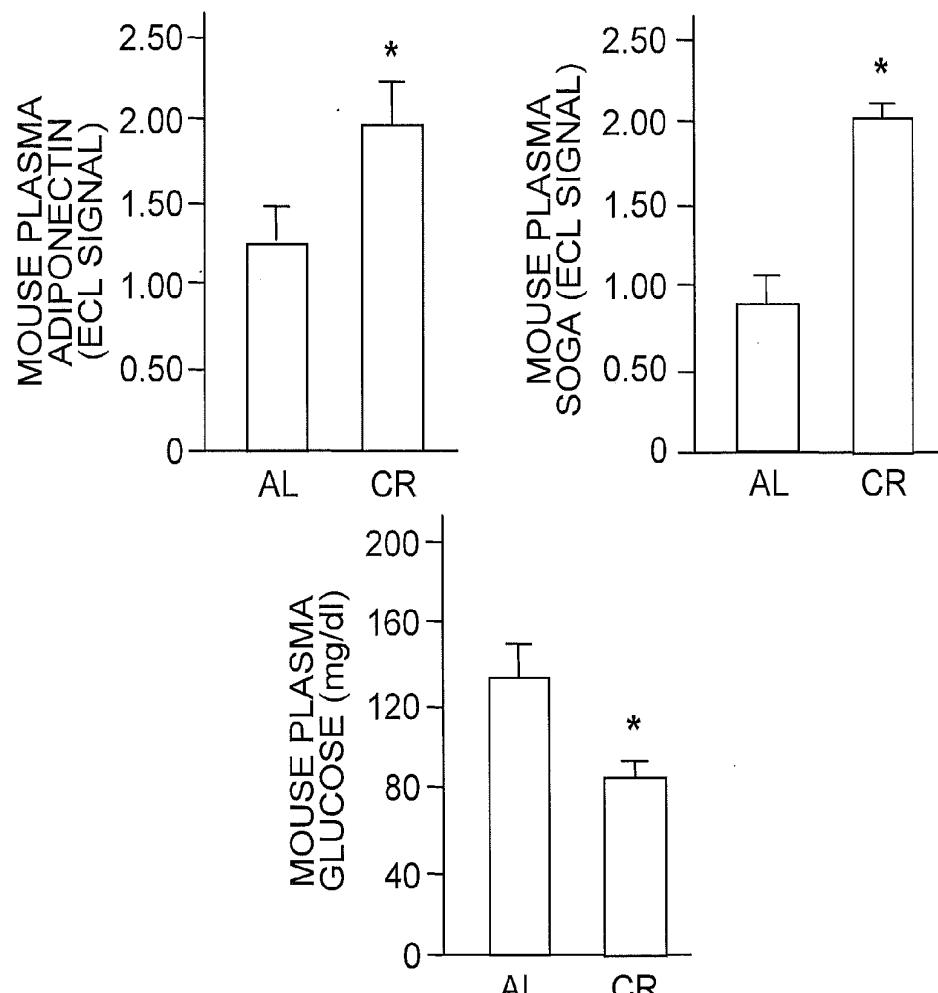


FIG. 14B

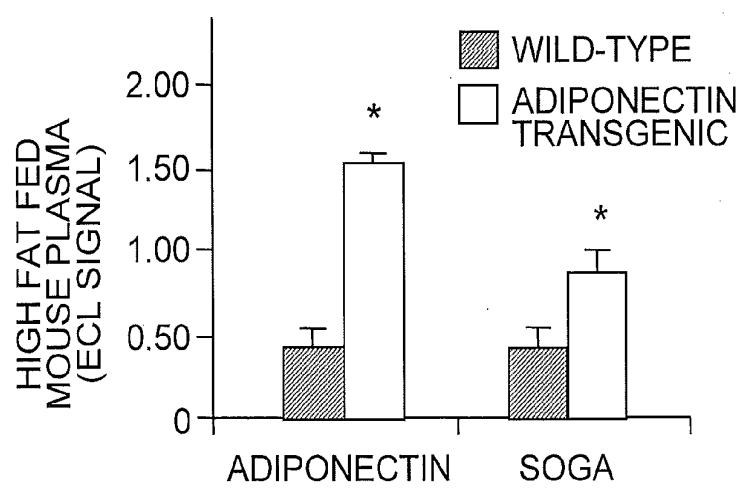


FIG. 14D

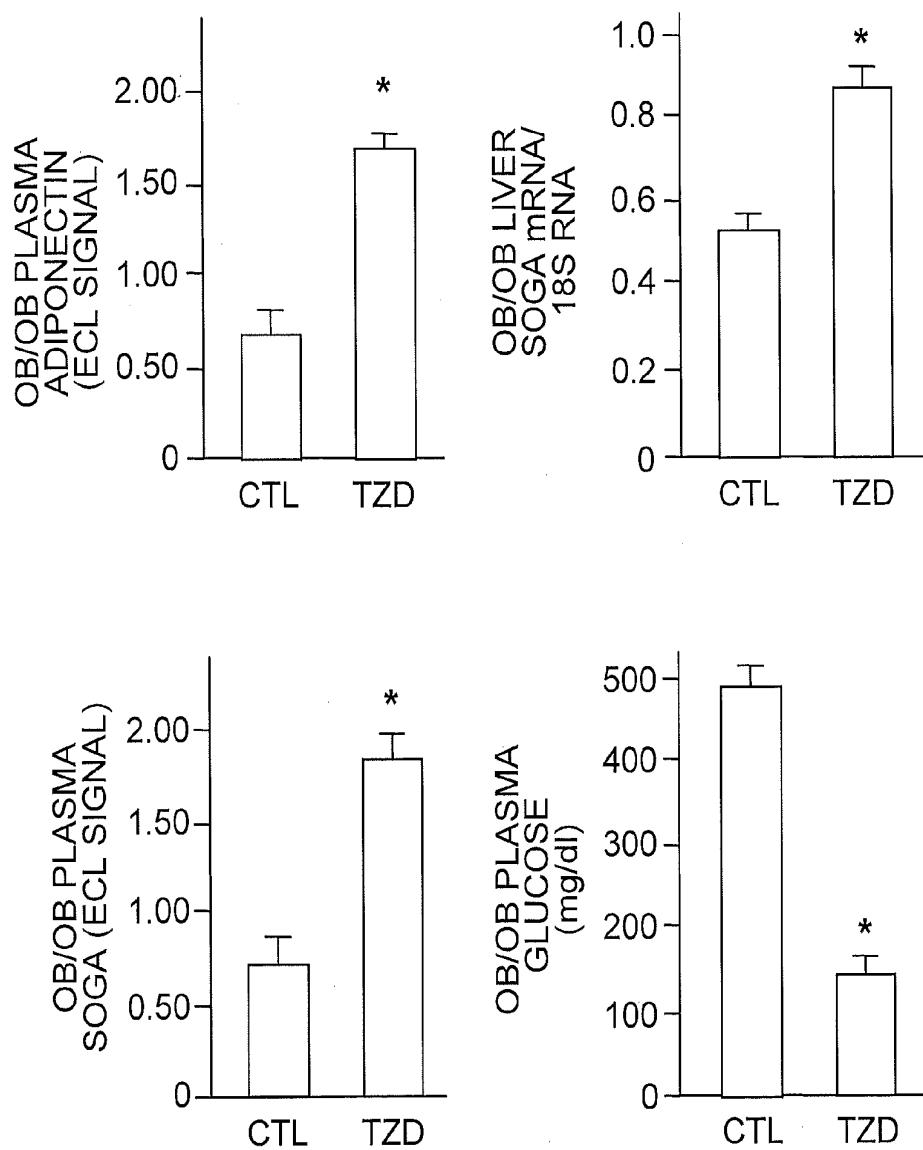


FIG. 14C

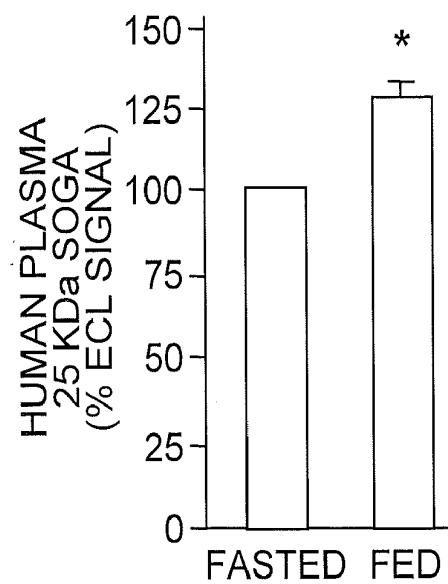


FIG. 15A

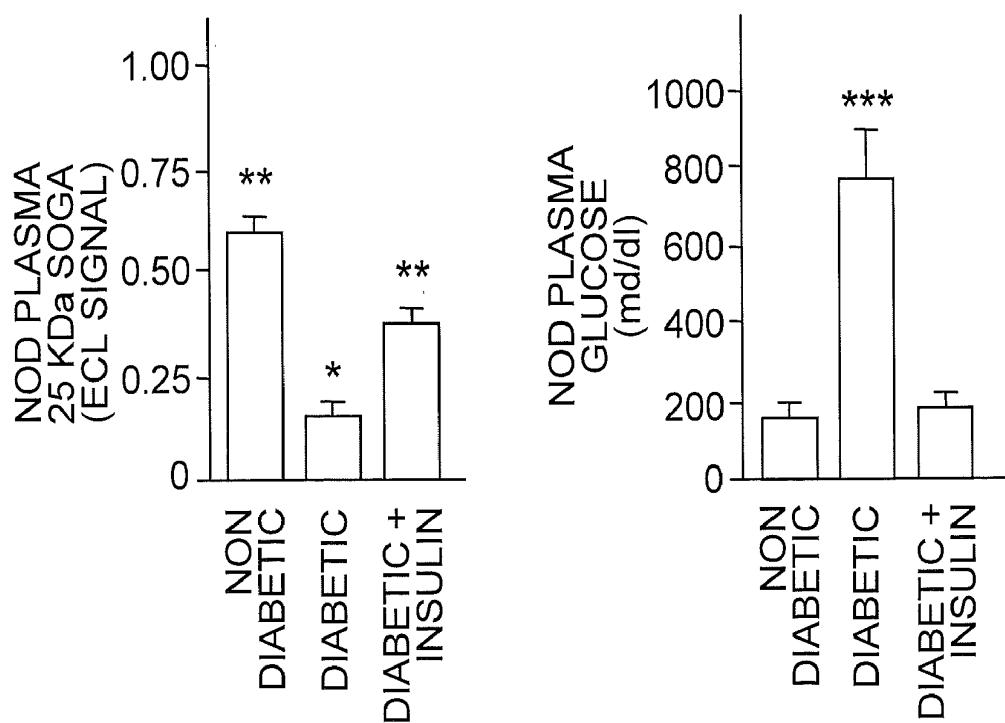


FIG. 15B

SOGA POLYNUCLEOTIDES AND POLYPEPTIDES AND USES THEREOF

STATEMENT OF FEDERAL SUPPORT

[0001] This invention was made, in part, with government support under grant numbers DK075573, DK056350, and ES010126 from the National Institutes of Health. The United States government has certain rights to this invention.

FIELD OF THE INVENTION

[0002] The present invention relates to the identification of polynucleotides and polypeptides involved in insulin and adiponectin signaling and regulation of glucose production. The invention further relates to the use of the identified polynucleotides and polypeptides, and inhibitors of the polynucleotides and polypeptides, in the regulation of glucose production and the monitoring and treatment of metabolic disorders such as diabetes.

BACKGROUND OF THE INVENTION

[0003] Adipose tissue exerts a powerful effect on glucose metabolism by regulating the concentration of circulating adiponectin (Goldline et al., *Lancet* 362:1431 (2003)). High adiponectin in the lean state is linked to elevated insulin sensitivity whereas low adiponectin in the obese state is linked to insulin resistance and diabetes (Arita et al., *Biochem. Biophys. Res. Commun.* 257:79 (1999); Hotta et al., *Arterioscler. Thromb. Vasc. Biol.* 20:1595 (2000); Maeda et al., *Diabetes* 50:2094 (2001); Weyer et al., *J. Clin. Endocrinol. Metab.* 2001, 86:1930 (2001)). Endogenous glucose production is elevated in diabetes (Wahren et al., *Annu. Rev. Nutr.* 27:329 (2007)). Studies in mice and liver cells show that adiponectin lowers glucose production by increasing the insulin sensitivity of the liver (Berg et al., *Nat. Med.* 7:947 (2001); Combs et al., *J. Clin. Invest.* 108:1875 (2001); Combs et al., *Endocrinology* 145:367 (2004)).

[0004] The signal transduction pathway of adiponectin is currently linked to (a) adiponectin receptors that bind to the full-length or the carboxy-terminal 'globular' fragment of adiponectin, (b) binding of the intracellular domains of adiponectin receptors 1 and 2 to the adaptor APPL1 and (c) the activation of AMPK, a signaling intermediate that reduces the gene expression of rate limiting enzymes for glucose production (Combs et al., *J. Clin. Invest.* 108:1875-(2001); Combs et al., *Endocrinology* 145:367 (2004); Tomas et al., *Proc. Natl. Acad. Sci. USA* 99:16309 (2002); Yamauchi et al., *Nat. Med.* 8:1288 (2002); Shklyav et al., *Proc. Natl. Acad. Sci. USA* 100:14217 (2003); Nawrocki et al., *J. Biol. Chem.* 281:2654 (2006); Andreelli et al., *Endocrinology* 147:2432 (2006); Mao et al., *Nat. Cell Biol.* 8:516 (2006); Brooks et al., *J. Biol. Chem.* 282:35069 (2007); Yoon et al., *Exp. Mol. Med.* 41:577 (2009); Wang et al., *J. Biol. Chem.* 282:7991 (2007)). However, the inhibition of glucose production by this pathway is not completely clear.

[0005] Glucose production depends on autophagy, a regulated mechanism of intracellular degradation that is inhibited by insulin (Amherdt et al., *J. Clin. Invest.* 54:188 (1974)). Autophagy provides the biochemical intermediates for glucose production through the hydrolysis of proteins, glycogen and triglycerides (Mortimore et al., *Annu. Rev. Nutr.* 7:539 (1987); Kotoulas et al., *Pathol. Res. Pract.* 202:631 (2006); Singh et al., *Nature* 458:1131 (2009)). Insulin inhibition of autophagy in isolated hepatocytes is linked to the activation of

mTOR (Blommaart et al., *J. Biol. Chem.* 270:2320 (1995); Kanazawa et al., *J. Biol. Chem.* 279:8452 (2004)). Hence, reports that AMPK, an essential mediator of adiponectin action, inhibits mTOR and stimulates autophagy are perplexing (Shaw et al., *Cancer Cell* 6:91 (2004); Meley et al., *J. Biol. Chem.* 281:34870 (2006); Xu et al., *Cell Death Differ.* 14:1948 (2007); Liang et al., *Nat. Cell Biol.* 9:218-(2007); Meijer et al., *Autophagy* 3:238 (2007); Cheng et al., *J. Biol. Chem.* 279:15719 (2004); Hoyer-Hansen et al., *Mol. Cell.* 25:193 (2007)).

[0006] The present invention addresses previous shortcomings in the art by providing a novel polynucleotide and polypeptide that connects insulin, adiponectin, and glucose production and that can be used for diagnostic and therapeutic methods.

SUMMARY OF THE INVENTION

[0007] The present invention is based, in part, on the identification of a novel polypeptide named Suppressor of Glucose by Autophagy (SOGA), also known as Target of Adiponectin (TOA), and the role it plays in insulin and adiponectin signaling and glucose production. The invention is based further on the use of this polypeptide, polynucleotides encoding the polypeptide, and inhibitors thereof, in the regulation of glucose production and the monitoring and treatment of metabolic disorders related to glucose levels, such as diabetes.

[0008] Accordingly, as one aspect, the invention provides an isolated polynucleotide selected from the group consisting of:

[0009] (a) a polynucleotide comprising a nucleotide sequence at least 70% (e.g., 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100%) identical to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3 and encoding a functional SOGA polypeptide;

[0010] (b) a polynucleotide that hybridizes to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3 under stringent hybridization conditions and encodes a functional SOGA polypeptide;

[0011] (c) a polynucleotide encoding a functional SOGA polypeptide comprising an amino acid sequence at least 70% (e.g., 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100%) identical to an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 4; and

[0012] (d) a functional fragment of any of (a) to (c).

[0013] The invention further relates to vectors and cells comprising the polynucleotides of the invention, and methods of recombinantly expressing the polypeptides of the invention.

[0014] Another aspect of the invention relates to isolated SOGA polypeptides or functional fragments thereof encoded by the isolated polynucleotides of the invention. Functional fragments include, without limitation, C-terminal fragments of about 80 kDa and about 25 kDa. In some embodiments, the polypeptide is part of a fusion protein.

[0015] A further aspect of the invention relates to agents that inhibit the expression and/or activity of SOGA polypeptides or polynucleotides, including antibodies, antisense oligonucleotides, ribozymes, siRNAs, and small molecules.

[0016] An additional aspect of the invention relates to pharmaceutical compositions comprising the polypeptides, polynucleotides, or inhibitory agents of the invention.

[0017] A further aspect of the invention relates to non-human animals genetically modified to express the polypeptide of the invention or to inhibit expression of the polypeptide of the invention.

[0018] Another aspect of the invention relates to methods of decreasing glucose production in a cell or decreasing autophagy in a cell, comprising contacting the cell with the polypeptides or polynucleotides of the invention.

[0019] A further aspect of the invention relates to methods of decreasing blood glucose levels in a subject or of increasing insulin sensitivity in a subject, comprising delivering to the subject the polypeptides or polynucleotides of the invention.

[0020] Another aspect of the invention relates to methods of increasing glucose production in a cell or increasing autophagy in a cell, comprising contacting the cell with an agent that decreases the expression and/or activity of the polypeptides or polynucleotides of the invention.

[0021] Another aspect of the invention relates to methods of increasing blood glucose levels in a subject or of decreasing insulin sensitivity in a subject, comprising delivering to the subject an agent that decreases the activity of the polypeptides or polynucleotides of the invention.

[0022] An additional aspect of the invention relates to a method of measuring the response of a subject to a treatment for diabetes, comprising determining the circulating level of the polypeptides of the invention in the subject after administration of the treatment and comparing it to the circulating level of the polypeptide in the subject before administration of the treatment.

[0023] Another aspect of the invention relates to a method of predicting the clinical outcome of a diabetes treatment in a subject, comprising determining the circulating level of the polypeptide of the invention in the subject after administration of the treatment and comparing it to the circulating level of the polypeptide in the subject before administration of the treatment.

[0024] Another aspect of the invention relates to a method of identifying an agent that binds to the polypeptides of the invention, comprising contacting the polypeptide or a functional fragment thereof with a test agent under conditions whereby binding between the polypeptide or a functional fragment thereof and the test agent can occur; and detecting binding between the polypeptide or a functional fragment thereof and the test agent.

[0025] An additional aspect of the invention relates to a method of identifying an agent that modulates the activity of polypeptides of the invention, comprising contacting the polypeptide or a functional fragment thereof with a test agent under conditions whereby modulation of the activity of the polypeptide or a functional fragment thereof can occur; and detecting modulation of the activity of the polypeptide or a functional fragment thereof upon contact with the test agent as compared to activity of the polypeptide or a functional fragment thereof in the absence of contact with the test agent.

[0026] A further aspect of the invention relates to a kit comprising a reagent for determining the expression and/or activity of the polypeptides and/or polynucleotide of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] FIG. 1 shows an amino acid sequence analysis of SOGA for conserved functional domains.

[0028] FIG. 2 shows the current model of autophagocytosis and the autophagy machinery showing mTOR and ATG 16 in black.

[0029] FIG. 3 shows proteolytic cleavage of SOGA yielding a circulating 25 kDa C-terminal fragment.

[0030] FIG. 4 shows that antisera from two different rabbits immunized with two different peptide antigens, 476 and 477, detected a 25 kDa band in mouse plasma.

[0031] FIG. 5 shows that the concentration of SOGA in plasma corresponded with circulating levels of adiponectin.

[0032] FIG. 6 shows western blot and densitometry of adiponectin and SOGA in ob/ob control mice and ob/ob mice treated with pioglitazone.

[0033] FIG. 7 shows western blot and densitometry of adiponectin and SOGA in ad libitum and calorie restricted fed C57B1 mice.

[0034] FIG. 8 shows western blot and densitometry of adiponectin and SOGA in rapamycin and control fed C57B1 mice.

[0035] FIG. 9 shows FPLC fraction analysis of mouse plasma for SOGA.

[0036] FIGS. 10A-10B show the sequence (SEQ ID NO:2) and predicted functional domains of SOGA.

[0037] FIGS. 11A-11D show the function and regulation of SOGA in primary hepatocytes.

[0038] FIGS. 12A-12C show detection of circulating SOGA in mice.

[0039] FIGS. 13A-13B show detection of recombinant SOGA.

[0040] FIGS. 14A-14D show the circulating levels of adiponectin and SOGA in humans and mice.

[0041] FIGS. 15A-15B show the circulating levels of SOGA in relation to insulin in humans and mice.

DETAILED DESCRIPTION OF THE INVENTION

[0042] The present invention will now be described in more detail with reference to the accompanying drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art.

[0043] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The terminology used in the description of the invention herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. All publications, patent applications, patents, patent publications and other references cited herein are incorporated by reference in their entireties for the teachings relevant to the sentence and/or paragraph in which the reference is presented.

[0044] Unless the context indicates otherwise, it is specifically intended that the various features of the invention described herein can be used in any combination.

[0045] Moreover, the present invention also contemplates that in some embodiments of the invention, any feature or combination of features set forth herein can be excluded or omitted.

[0046] Nucleotide sequences are presented herein by single strand only, in the 5' to 3' direction, from left to right, unless specifically indicated otherwise. Nucleotides and amino

acids are represented herein in the manner recommended by the IUPAC-IUB Biochemical Nomenclature Commission, or (for amino acids) by either the one-letter code, or the three letter code, both in accordance with 37 C.F.R. §1.822 and established usage.

[0047] Except as otherwise indicated, standard methods known to those skilled in the art may be used for cloning genes, amplifying and detecting nucleic acids, and the like. Such techniques are known to those skilled in the art. See, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual 2nd Ed. (Cold Spring Harbor, N.Y., 1989); Ausubel et al. Current Protocols in Molecular Biology (Green Publishing Associates, Inc. and John Wiley & Sons, Inc., New York).

I. Definitions

[0048] As used in the description of the invention and the appended claims, the singular forms "a," "an," and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise.

[0049] Also as used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative ("or").

[0050] The term "about," as used herein when referring to a measurable value such as an amount of polypeptide, dose, time, temperature, enzymatic activity or other biological activity and the like, is meant to encompass variations of $\pm 20\%$, $\pm 10\%$, $\pm 5\%$, $\pm 1\%$, $\pm 0.5\%$, or even $\pm 0.1\%$ of the specified amount.

[0051] The term "consists essentially of" (and grammatical variants), as applied to a polynucleotide or polypeptide sequence of this invention, means a polynucleotide or polypeptide that consists of both the recited sequence (e.g., SEQ ID NO) and a total of ten or less (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) additional nucleotides or amino acids on the 5' and/or 3' or N-terminal and/or C-terminal ends of the recited sequence such that the function of the polynucleotide or polypeptide is not materially altered. The total of ten or less additional nucleotides or amino acids includes the total number of additional nucleotides or amino acids on both ends added together. The term "materially altered," as applied to polynucleotides of the invention, refers to an increase or decrease in ability to express the encoded polypeptide of at least about 50% or more as compared to the expression level of a polynucleotide consisting of the recited sequence. The term "materially altered," as applied to polypeptides of the invention, refers to an increase or decrease in the ability to inhibit glucose production of at least about 50% or more as compared to the activity of a polypeptide consisting of the recited sequence.

[0052] A "therapeutically effective" amount as used herein is an amount that provides some improvement or benefit to the subject. Alternatively stated, a "therapeutically effective" amount is an amount that will provide some alleviation, mitigation, or decrease in at least one clinical symptom in the subject (e.g., in the case of diabetes, reduction in glucose levels or increase in insulin sensitivity). Those skilled in the art will appreciate that the therapeutic effects need not be complete or curative, as long as some benefit is provided to the subject.

[0053] By the terms "treat," "treating," or "treatment of," it is intended that the severity of the subject's condition is

reduced or at least partially improved or modified and that some alleviation, mitigation or decrease in at least one clinical symptom is achieved.

[0054] The term "control sample," as used herein, refers to a tissue or cell sample that is used to compare the level of expression and/or activity of a SOGA polypeptide to the level of expression and/or activity in a sample of interest. The control sample may be, for example, from a normal (i.e., non-diseased) portion of the same tissue or cell type in the subject, from a different tissue or cell type in the subject, from a matched individual, or may be a standard derived from the average of measurements taken from a population of subjects. In another embodiment, the control sample may be from the disease tissue of the subject, e.g., at the time of diagnosis, prior to treatment, or after a stage of treatment.

[0055] As used herein, "nucleic acid," "nucleotide sequence," and "polynucleotide" are used interchangeably and encompass both RNA and DNA, including cDNA, genomic DNA, mRNA, synthetic (e.g., chemically synthesized) DNA or RNA and chimeras of RNA and DNA. The term polynucleotide, nucleotide sequence, or nucleic acid refers to a chain of nucleotides without regard to length of the chain. The nucleic acid can be double-stranded or single-stranded. Where single-stranded, the nucleic acid can be a sense strand or an antisense strand. The nucleic acid can be synthesized using oligonucleotide analogs or derivatives (e.g., inosine or phosphorothioate nucleotides). Such oligonucleotides can be used, for example, to prepare nucleic acids that have altered base-pairing abilities or increased resistance to nucleases. The present invention further provides a nucleic acid that is the complement (which can be either a full complement or a partial complement) of a nucleic acid, nucleotide sequence, or polynucleotide of this invention.

[0056] An "isolated polynucleotide" is a nucleotide sequence (e.g., DNA or RNA) that is not immediately contiguous with nucleotide sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. Thus, in one embodiment, an isolated nucleic acid includes some or all of the 5' non-coding (e.g., promoter) sequences that are immediately contiguous to a coding sequence. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment), independent of other sequences. It also includes a recombinant DNA that is part of a hybrid nucleic acid encoding an additional polypeptide or peptide sequence. An isolated polynucleotide that includes a gene is not a fragment of a chromosome that includes such gene, but rather includes the coding region and regulatory regions associated with the gene, but no additional genes naturally found on the chromosome.

[0057] The term "isolated" can refer to a nucleic acid or polypeptide that is substantially free of cellular material, viral material, and/or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an "isolated fragment" is a fragment of a nucleic acid, nucleotide sequence or polypeptide that is not naturally occurring as a fragment and would not be found in the natural state. "Isolated" does not mean that the preparation is technically pure (homogeneous), but it is sufficiently pure to provide the

polypeptide or nucleic acid in a form in which it can be used for the intended purpose. In certain embodiments, the polypeptide is at least about 50% pure, e.g., at least about 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% or more pure.

[0058] An isolated cell refers to a cell that is separated from other components with which it is normally associated in its natural state. For example, an isolated cell can be a cell in culture medium and/or a cell in a pharmaceutically acceptable carrier of this invention. Thus, an isolated cell can be delivered to and/or introduced into a subject. In some embodiments, an isolated cell can be a cell that is removed from a subject and manipulated as described herein *ex vivo* and then returned to the subject.

[0059] The term "fragment," as applied to a polynucleotide, will be understood to mean a nucleotide sequence of reduced length relative to a reference nucleic acid or nucleotide sequence and comprising, consisting essentially of, and/or consisting of a nucleotide sequence of contiguous nucleotides identical or almost identical (e.g., 90%, 92%, 95%, 98%, 99% identical) to the reference nucleic acid or nucleotide sequence. Such a nucleic acid fragment according to the invention may be, where appropriate, included in a larger polynucleotide of which it is a constituent. In some embodiments, such fragments can comprise, consist essentially of, and/or consist of oligonucleotides having a length of at least about 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 400, 500, or more consecutive nucleotides of a nucleic acid according to the invention. In some embodiments, such fragments can comprise, consist essentially of, and/or consist of oligonucleotides having a length of less than about 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 400, or 500 consecutive nucleotides of a nucleic acid according to the invention.

[0060] The term "fragment," as applied to a polypeptide, will be understood to mean an amino acid sequence of reduced length relative to a reference polypeptide or amino acid sequence and comprising, consisting essentially of, and/or consisting of an amino acid sequence of contiguous amino acids identical or almost identical (e.g., 90%, 92%, 95%, 98%, 99% identical) to the reference polypeptide or amino acid sequence. Such a polypeptide fragment according to the invention may be, where appropriate, included in a larger polypeptide of which it is a constituent. In some embodiments, such fragments can comprise, consist essentially of, and/or consist of peptides having a length of at least about 4, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 300, 400, 500, or more consecutive amino acids of a polypeptide or amino acid sequence according to the invention. In some embodiments, such fragments can comprise, consist essentially of, and/or consist of peptides having a length of less than about 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 400, or 500 consecutive nucleotides of a nucleic acid according to the invention.

[0061] The term "functional SOGA polypeptide," as applied herein, refers to a polypeptide that substantially retains at least one biological activity normally associated with the naturally occurring SOGA polypeptide (e.g., the ability to inhibit glucose production, protein binding, ligand or receptor binding). In particular embodiments, the "functional" polypeptide substantially retains all of the activities possessed by the naturally occurring polypeptide. By "substantially retains" biological activity, it is meant that the polypeptide retains at least about 20%, 30%, 40%, 50%, 60%,

75%, 85%, 90%, 95%, 97%, 98%, 99%, or more, of the biological activity of the native polypeptide (and can even have a higher level of activity than the native polypeptide). A "non-functional" polypeptide is one that exhibits little or essentially no detectable biological activity normally associated with the polypeptide (e.g., at most, only an insignificant amount, e.g., less than about 10% or even 5%). Biological activities such as protein binding and suppression of glucose production can be measured using assays that are well known in the art and as described herein. In certain embodiments, the "activity" of a SOGA polypeptide is defined as the ability to inhibit glucose production in a population of isolated hepatocytes (either primary hepatocytes or a hepatocyte cell line).

[0062] The term "functional fragment," as applied to a polypeptide, refers to a fragment that substantially retains at least one biological activity of the full length polypeptide, e.g., the ability to inhibit glucose production. By "substantially retains" biological activity, it is meant that the fragment retains at least about 20%, 30%, 40%, 50%, 60%, 75%, 85%, 90%, 95%, 97%, 98%, 99%, or more, of the biological activity of the full length polypeptide (and can even have a higher level of activity than the full length polypeptide). A "non-functional" fragment is one that exhibits little or essentially no detectable biological activity normally associated with the polypeptide (e.g., at most, only an insignificant amount, e.g., less than about 10% or even 5%).

[0063] The term "functional fragment," as applied to a polynucleotide, refers to a polynucleotide that encodes a functional fragment of a polypeptide.

[0064] A "vector" is any nucleic acid molecule for the cloning of and/or transfer of a nucleic acid into a cell. A vector may be a replicon to which another nucleotide sequence may be attached to allow for replication of the attached nucleotide sequence. A "replicon" can be any genetic element (e.g., plasmid, phage, cosmid, chromosome, viral genome) that functions as an autonomous unit of nucleic acid replication *in vivo*, i.e., capable of replication under its own control. The term "vector" includes both viral and nonviral (e.g., plasmid) nucleic acid molecules for introducing a nucleic acid into a cell *in vitro*, *ex vivo*, and/or *in vivo*. A large number of vectors known in the art may be used to manipulate nucleic acids, incorporate response elements and promoters into genes, etc. For example, the insertion of the nucleic acid fragments corresponding to response elements and promoters into a suitable vector can be accomplished by ligating the appropriate nucleic acid fragments into a chosen vector that has complementary cohesive termini. Alternatively, the ends of the nucleic acid molecules may be enzymatically modified or any site may be produced by ligating nucleotide sequences (linkers) to the nucleic acid termini. Such vectors may be engineered to contain sequences encoding selectable markers that provide for the selection of cells that contain the vector and/or have incorporated the nucleic acid of the vector into the cellular genome. Such markers allow identification and/or selection of host cells that incorporate and express the proteins encoded by the marker. A "recombinant" vector refers to a viral or non-viral vector that comprises one or more heterologous nucleotide sequences (i.e., transgenes), e.g., two, three, four, five or more heterologous nucleotide sequences.

[0065] Viral vectors have been used in a wide variety of gene delivery applications in cells, as well as living animal subjects. Viral vectors that can be used include, but are not limited to, retrovirus, lentivirus, adeno-associated virus, poxvirus, alphavirus, baculovirus, vaccinia virus, herpes virus,

Epstein-Barr virus, and adenovirus vectors. Non-viral vectors include plasmids, liposomes, electrically charged lipids (cytosefctins), nucleic acid-protein complexes, and biopolymers. In addition to a nucleic acid of interest, a vector may also comprise one or more regulatory regions, and/or selectable markers useful in selecting, measuring, and monitoring nucleic acid transfer results (delivery to specific tissues, duration of expression, etc.).

[0066] Vectors may be introduced into the desired cells by methods known in the art, e.g., transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (lysosome fusion), use of a gene gun, or a nucleic acid vector transporter (see, e.g., Wu et al., *J. Biol. Chem.* 267:963 (1992); Wu et al., *J. Biol. Chem.* 263:14621 (1988); and Hartmut et al., *Canadian Patent Application No. 2,012,311*, filed Mar. 15, 1990).

[0067] In some embodiments, a polynucleotide of this invention can be delivered to a cell *in vivo* by lipofection. Synthetic cationic lipids designed to limit the difficulties and dangers encountered with liposome-mediated transfection can be used to prepare liposomes for *in vivo* transfection of a nucleotide sequence of this invention (Feigner et al., *Proc. Natl. Acad. Sci. USA* 84:7413 (1987); Mackey, et al., *Proc. Natl. Acad. Sci. USA* 85:8027 (1988); and Ulmer et al., *Science* 259:1745 (1993)). The use of cationic lipids may promote encapsulation of negatively charged nucleic acids, and also promote fusion with negatively charged cell membranes (Feigner et al., *Science* 337:387 (1989)). Particularly useful lipid compounds and compositions for transfer of nucleic acids are described in International Patent Publications WO95/18863 and WO96/17823, and in U.S. Pat. No. 5,459,127. The use of lipofection to introduce exogenous nucleotide sequences into specific organs *in vivo* has certain practical advantages. Molecular targeting of liposomes to specific cells represents one area of benefit. It is clear that directing transfection to particular cell types would be particularly preferred in a tissue with cellular heterogeneity, such as pancreas, liver, kidney, and the brain. Lipids may be chemically coupled to other molecules for the purpose of targeting (Mackey, et al., 1988, *supra*). Targeted peptides, e.g., hormones or neurotransmitters, and proteins such as antibodies, or non-peptide molecules can be coupled to liposomes chemically.

[0068] In various embodiments, other molecules can be used for facilitating delivery of a nucleic acid *in vivo*, such as a cationic oligopeptide (e.g., WO95/21931), peptides derived from nucleic acid binding proteins (e.g., WO96/25508), and/or a cationic polymer (e.g., WO95/21931).

[0069] It is also possible to introduce a vector *in vivo* as naked nucleic acid (see U.S. Pat. Nos. 5,693,622, 5,589,466

and 5,580,859). Receptor-mediated nucleic acid delivery approaches can also be used (Curiel et al., *Hum. Gene Ther.* 3:147 (1992); Wu et al., *J. Biol. Chem.* 262:4429 (1987)).

[0070] The term "transfection" or "transduction" means the uptake of exogenous or heterologous nucleic acid (RNA and/or DNA) by a cell. A cell has been "transfected" or "transduced" with an exogenous or heterologous nucleic acid when such nucleic acid has been introduced or delivered inside the cell. A cell has been "transformed" by exogenous or heterologous nucleic acid when the transfected or transduced nucleic acid imparts a phenotypic change in the cell and/or a change in an activity or function of the cell. The transforming nucleic acid can be integrated (covalently linked) into chromosomal DNA making up the genome of the cell or it can be present as a stable plasmid.

[0071] As used herein, the terms "protein" and "polypeptide" are used interchangeably and encompass both peptides and proteins, unless indicated otherwise.

[0072] A "fusion protein" is a polypeptide produced when two heterologous nucleotide sequences or fragments thereof coding for two (or more) different polypeptides not found fused together in nature are fused together in the correct translational reading frame. Illustrative fusion polypeptides include fusions of a polypeptide of the invention (or a fragment thereof) to all or a portion of glutathione-S-transferase, maltose-binding protein, or a reporter protein (e.g., Green Fluorescent Protein, β -glucuronidase, β -galactosidase, luciferase, etc.), hemagglutinin, c-myc, FLAG epitope, etc.

[0073] By the term "express" or "expression" of a polynucleotide coding sequence, it is meant that the sequence is transcribed, and optionally, translated. Typically, according to the present invention, expression of a coding sequence of the invention will result in production of the polypeptide of the invention. The entire expressed polypeptide or fragment can also function in intact cells without purification.

II. SOGA Polynucleotides and Polypeptides

[0074] In one aspect, the invention relates to an isolated polynucleotide encoding a SOGA polypeptide or a functional fragment thereof. In one embodiment, the SOGA polypeptide is a mammalian SOGA polypeptide, e.g., human or mouse. The cDNA, polypeptide, and genomic sequences of mouse SOGA have been deposited in GenBank under Accession No. H977045 and are disclosed herein as SEQ ID NOS:1, 2, and 10, respectively. The cDNA, polypeptide, and genomic sequences of human SOGA are disclosed herein as SEQ ID NOS:3, 4, and 11, respectively. The polynucleotide can comprise cDNA sequences, genomic sequences, synthetic sequences, or combinations thereof.

Mouse SOGA cDNA Sequence

(SEQ ID NO: 1)

```

atgtgggcctggagctggcgctgagcagcgacgcccagtcggccggccggccggccgg 60
ggaccggcaccgggcagccgcggccagccagcgcagtcgggcagcagcctccggggccc 120
ccgcattcccgatgagccgtcggtggccgatgtcggtggcagcagccgttgccat 180
tcagegcctcgtagccttctccgacccatccgaggagatgctggactgtggggccggag 240
gctttgggtgcgggagctgaaagagctgcgttccgagaacgactatctcaaggatgagat 300
aggagactacggcgtgagatgctggagatgcggatgtctacatggaggaagacgtgtatc 360

```

-continued

agctgcagtacccgactcgctgaaggctggcccgccagcctccgcgtgcccagacaggcc 420
aggttcatggggactcatccgaggctggaaacaggacgtcaaggctctaaggacatct 480
ccatgcggcttcacaaggagctggagggtggatggagaagaagcggataggctggaggagg 540
agaacgaggggcttcgacagaggcttattgagacagagactggccaagcagggtctacaga 600
cggagctggatcgccccagagcattcttgaagaaaagagaaaccggctctgggaa 660
agacagataagaaggctactgcacaggaggatagtgcagacactgaagtgcagctgcatt 720
ttgcaaaggaggagtggccctcatgtgcagaagctcaccaagttggctaaggagaacg 780
acagcatgaaggaggactctcaagtacagatcgcttatgggacactggatgcagccc 840
tgtcgccagaggactggccgtgctccgcactccgtgagactgagctgaagggtgcacc 900
tgaagctggtgaggaggaggccaaacctgtgagccggcgatagtggagctggagggtgg 960
agaaccgtggcctgcgagccgatggacatgaaggaccacgggggtggcgggggg 1020
ccgaggccaggctggccctctttctgggtggtagtgcggggagagcctagccgagt 1080
tgcggccacactgcagttgtggaaagaggaggactgactgtgaggcgctcctcagctg 1140
agctggaggaccagaacaagttgtctgtgcacgcgactggccaaataccgtcgagcag 1200
agctggacgtgacgctgtcgaggacagactgcctgtcagcgagccctcgaggagg 1260
agctggcagcccaagctgcagatcgccgagactcgccgcaaggtaagaagctgcagt 1320
atgagaaccgcgtgtcccttccaatctgcagcgctgtgacctggccctgcagagca 1380
cacgcggccatgtggagacggacgctggggactctgcgcgtgcgtgcctgcctggcc 1440
ctctgggtgagacgctggagccccacggccggctgtgcaggggccgtgaagecgagg 1500
cgctggccggctacggggcaggccgttggtagcaggccatcgacgtcctgggtgg 1560
ctgatgcctatggcttcagtcggccctccgcctgtgcctggacaatgagtgtgtact T620
tgcactgcacggcgcctgacaacacgcgaggggccaggatgccaagctcatccacg 1680
ccatctggtgccgtgactgttttgcacggcccttcggccctgtgcctggccctgtgcct 1740
atgtggccctggggagctctggcaaggagcagccgtgagcccttcctgtgcctgcct 1800
tgggctcccaggccctgtcaaggagatcacgtgtccaaagacccctggctctgacttcc 1860
agccacctgacttcagagacctgttgagtgggagccaggatccgagaggccctccgta 1920
ccggggacttggagttcaagccctgacccttagtcggaaacttcaggccctaccgagctgaag 1980
ataacgattctatgcctctgagatcaaggatcttcagctggccctggccgaggccacg 2040
acagccctccggggcttgcaagacgcagactgtcccgaggacggcagctccggaaaggaggagg 2100
ctgacagcttcaaccagaaaaatyguccagctgaaggaaagccaggcagaggccgtctga 2160
gacgggagtttagctgcagactgtgagcttcggccctggccgactggagcagaaggcttgc 2220
gccaagagaagaacatctggcaggagtcggccatgtcaagcacaacttctgtgc 2280
tcttcatgaagctccgggtggcttcgtggccgtggccaggccatgtgcctggcc 2340
aagaggatgacttcctggaggtgaacagcatgaaggactgtacctgtatggaggaag 2400
aggagatgaacgcacactcggtataacaaggccctgcacaggggagagctggacccaga 2460
acacgcctaatgagtgcataagacccctggccgacatgaaggctcaccctgaaggagctgt 2520
gctggctgtccaggacgcgtcggtgtactgaacttcagcagcagatgcgaaagg 2580
ccaaggccacctggagacagacgcgtgcagactcaagggccacgcctcgagatggagc 2640
tgaaggctggaaagggtgccagtgagaggccggccctgacttgcactgcactgcaga 2700

-continued

gagagcgggaggaggcagcaacacacctctggcagagtccatcagcgccgtcatggagctga 2760
 cgaggcagctcagctgagcgcggccactggagacaggagaagctcagctgggg 2820
 ggctcagggagaaaagcagcggctggagcagcggtaaggagctcagaaccgccta 2880
 gtcagttgcagaaggctgccgagccctgggtcctgaagcactcagacatggagaagcaag 2940
 acaacagctggaaagaggcacgaaagtggagaagacccatgacaaggagggtgtctctgaag 3000
 ctgagctcggggactggcutaaagaggaccataatcagtctcctccatgtctgagttt 3060
 aaagttgctcactgtccccgtaccttgcggggatgcccggaaacaagaagctgc 3120
 ccaacggccctgcgtttgccttgtagtactggccagtgaggccctgagaaagacgcca 3180
 aggagaaggcgggcttccaccggactgtagccacattggtagttggcctgtcagg 3240
 aacctgcagggagacagatgcgcgcagctacacggctccagacaagacgggatccgag 3300
 tctactatagtccgcagtggtcggtcggtcggtgtccctgtggccatgacaaggagg 3360
 gcaagatcattcattgagccaggcttcttcaactaccgccaagcccaaggagccagccg 3420
 aggctgacgggctggccgagagctctacagccggtggtttgcatttctccggcagc 3480
 ggctggatggaggatcggggccagcacctcggttccggacotgtttccccgcattgc 3540
 atgactttgagatgtcgcccaatgtggatgtggacatgaaaggatcaccaactgcgtgc 3600
 ggcaggccatgcgtccggctcttgcggagaggtaaagaacacatccagccagacgg 3660
 taggcgtggccaccgtggcaccacagattcggtcggtcgttgcggatggccattc 3720
 acccaccgcagcgcctccacagcaagagctggtcacccgcagctcctcgcttgc 3780
 ctgtgcgcagcaagcagatcttcctccctgacaaggccattctcgattgacggc 3840
 catgttgcgtcccaagatcggctcaccaagctccagagacgtccggatgtccagg 3900
 atagcaccaaggaccgcgcgttgcggatggacccatgtggccatgtggccct 3960
 gggctgcgtccaccacaccacacgggatagccctgtactggatggccatgtggcc 4020
 ctagectctcttagtggatggactctggggcaccggactgtgtggaaactggca 4080
 tgtctgaggccgaaccaaacctgagccctccaaagtatggcattttcaggatcc 4140
 ggaacgtgtggccggcaccggactgtccggatggatggatcctgtatccggatcc 4200
 aaccagagcccccttcgcgcagccactaccatcaacccggatggatccggatcc 4260
 acaagaaggcggccaaggcagggtggtagcgaagaggctcagaccaccatgtgtcc 4320
 tggggaggatggcatttggatggatggatggatcctgtatccggatgtggatcc 4380
 ccgtatgtgacttagcgtccaggacttgcgttgcgttgcgtccggatccggatcc 4440
 ccatccggactctccctccaaagtgcaggctgcaccctcagatcggctgggggggg 4500
 aggagagggcagctceccagtgtggatccctgagcaaccaagcaccaccaatggc 4560
 gaccgcggagatggcggatggcggatggatggatggatggatggatggatggc 4620
 agctgtgccactgccacagaagagcttacattaaggtaaagcaagggtgttgcac 4680
 tgctggcagtgcacccatgttccaggaaagaca 4716

Mouse SOGA Polypeptide Sequence

(SEQ ID NO: 2)

MLDCGPGGLVHKLEELRSENDYLKDEIEELRAEMLEHKDVYMEBDVYQLQYRLRKAERRS 60

LRAAQTGQVDGELIRGLEQDVVKSKDISMRLKKELEVVEKKMRLEENKGLRQRRIETE 120

LAKQVLQTELDRPREHSLKKRGTRSLGKTDKKPTAQEDSADLKQCLHKAKEESALMCKKL 180

- continued

-continued

ccCtctccgcgtgggtggcgaggagtgccccggagatgtggcagagctgcccacacc 1080
tgtcagtttgtcgaagaggaggcccgagctgcCggcgctctgtcccgagctcgaggacc 1140
agaacaagctgtgctgaacgagctggccaagtccgtcgaggcacagatggacgtgg 1200
cgctgtcgaggacagctgttctgtcagcaacccacaggaggatggggccg 1260
ccaagctgcagatggcgagctcagcggcaaggtcaagaagcLgcagtacgagaaccgcg 1320
tgctcttcacccatccagcgtgtgacctcgccctctgcccagatgcgcggccatgc 1380
tggagacggacgcccggaggccggactctgccagtgctgtgcctgtccctggcgaga 1440
cacacgacccacccggccgtccgactctgcagagccaggaggcggatgtgcctggc 1500
tgagagagcaggccgcctggtcagtaaggccatcgatgtcctggctgtatgccaatg 1560
gctccacggctggcctccggctgtgtcgacaaacgagatgtgtgcattccggctgc 1620
aggccccccgacaacacgcggggcccccaggacaccaagctcatccatgcccattgg 1680
gcctgagcgtgtcgcagcaggagctgaatgccttacgcggaggcagatgcagtcctg 1740
ggtgctctgtcaaggaacacgcggagtccttcatcactgcggcccttggctcccagg 1800
ggctctctaaggagattttctggcaaaagacccctggctcagacttcagccacctgact 1860
tcagggacctgggaatggggccaggatccgcggaggcttcaggccattccgcactgg 1920
actctaagccgaccccgccggagttcaggccattccgcactgtgaagacaatgattct 1980
atgcctctgagatcaaggagctgcagctgtgtgtggctgaggccacgcgcctccgg 2040
gcttgcagagcagctcccgaggagccgcagtcacgaaaggaggaggccgacaattca 2100
accagaaaatggccsgotgaaggaggaccagcagggcgctctgaggcggagtttgc 2160
agctgcagagctgagctccgcggaggctggagcagcggcggaggatggcggagaaga 2220
acatgcggcgtgcaggagttcccgacattcaaggccacttcgtgtcttcatgaagc 2280
tcagggtggctctcaagcgctggcgccaggccaaaggccccccgcggcggaggatgt 2340
tcctcgagggtgaacacgcataaggagctgtacttgctgtatggaggaagaggagataa 2400
ctcagcatttcataacaaggccctgcacggggacagctggaccacacgcggcaatg 2460
agtacatcaagacactggccgcacatgaagggtgacgcgtgaaggagctgtgtgtcc 2520
gggatgaacgcgtggctgcaggagttcagcaacagtttgcggccaggatcttc 2580
gggagacagacggggcagagctcaaggccatctcccgatggagctgaagacaggga 2640
agggggccggggagcggggcaggccccactggaaaggcagccctacagcggagcgtgagg 2700
agcagcagcaccccttagtgcgttgcacgcgtgtcatggagactcggcagctgc 2760
agaccagtgcgcactggggccaggaaaagctgcagctggggagcggctgcagggtg 2820
agaageagcagggtggagcagcagggtgaaggagctgcagaaccgcctaagccagtc 2880
aggctgcgcacccctgggtctgcacactcgagctggagatggagaagcaggacaac 2940
aggagacacgcagtgcgttgcacgcacaaggaggctgttccgaaagttgagcttgg 3000
gaaatggttaaagagaaccaatctgtttctccatgtctgagttgaaagttgcgtcg 3060
actgttcccttacctcgctggcgaggatgcggccggcaagaagatgcctaacaaccctg 3120
cctttggctttgtgagctccgcggggatcccgagaaagacaccaaggagaaggctg 3180
ggctctcgaggactgcacccacctgggtgcgcctggccaggaccccccaggaa 3240
ggcagatgcagcgcagctacacggctctgacaagaegggcatccgagttactatagtc 3300

-continued

ccccggtggcccgccctcgagtcctgtggatcacaaagagggcaagatcatta 3360
 tcgagcccgcttctttaccacagccaagccaaagagtccgcggaggctatggc 3420
 tggctgagctctatggtcgggtctgcacttcacggcagccctggacggag 3480
 gtcagcggcagccccctcgccggccggcttggttcccagccgcctgcatactttg 3540
 agatgtcaggcaacatgagtatgacatgaaggagatcaccaacttgtgcccaggca 3600
 tgcgcctccggctactggagagggaaagttaagagacatccacggccacacggctgg 3660
 ccagtgtggcacacagaccatccgacggtcagcgtggcctgcagaccgaccaccc 3720
 gcacgcgcctccatggcaaggctgttcaccccgacgtcttcgtctgtgcga 3780
 gcaaggacatctccctccctggacaagggtccattcgccatcgagccgcctgtctgt 3840
 ccccaagtatggctaccaaagctccagggcggtctgttccaagctggacacgcaga 3900
 aggaccgcagcgtggaaacctgcaccaggcaagcagaacggctggcctggcccgct 3960
 ccaccaccacgcgggacagccctgtattgagaaacatcaacgatggacttccaccc 4020
 tcagtgtggtgagactcaggagcacggagtcgtgtggaaacttaggcattgtgaga 4080
 cgccggccaagcccgagccctccaaagtacggcattgtcaggaattttccgtatgtt 4140
 gtggccgggacccgagccccacccatcagcaggaggggtgtggcctggcccgct 4200
 ccctctcccaagccagctaccatcagccagagggtgtggcctggcccgatctgt 4260
 cagccaagttggcagcgtgaggaggtcagactcaccatgtccccaggtggaaagg 4320
 atggtgCcctcgggacggagatggagccgtgtgttccaaatgaggacgtgtttgt 4380
 actgttagtaccctgttccatcaccctgtgttccatcgccatccgc 4440
 artctccctccaaagtgcaggctgcacccttcagactccagctgggggtggggagg 4500
 cactccccccagcgagtgacagagcagccaagctcccgctcaaccagccagccct 4560
 ggatagcagaagggaaaccagcagagcagacgagcggaggtgaggcgagggctgt 4620
 cattgcctggccctgggggacacgcgtgtatgccactgcagaatgcagcttcacatca 4680
 aggttaagccgggtctctgtggccctgggtggtagcttcactggggaggaaagg 4740
 cagttagtgaggagagaccaaaccctgggtttccaaatgcactccactggagatctgt 4800
 gagccgatccctgggtcctaagagagacgcctgtgttctgcccattgcaccccttg 4860
 ga 4862

Human SOGA Polypeptide Sequence

(SKQ ID NO: 4)

MLDCGPSGLVRELEELRSENDYLKDEIEBLRAEMLEMARDVYMKEDDVYQLQYRLHKAERRS 60
 LRAAQQTGQVDGELIRGLEQDVVKSKDISMRLHKELEVVEKKRARLEEEENEELRQRRIETE 120
 IAKVQVLOTELERPRKHSLKRGTRSLGKADKKTLVQEDSADLKQHLKAKEESALMCKKL 180
 TKLAKENDSMKEELLKYRSLYGDLDALSAAEELADAPHSRETELKVHLKLVEEEANLLSR 240
 RIVELEVENRGLRAEMDDMKDHGGCGGPEARLAFSALGGGECCGESLAELRRHLQFVEEE 300
 AELLRRSSAKLEDQNKLNLNEAKPRSCHRDLVALSEDSCSVLSEPSQKKIAAKLQIGE 360
 LSGKVKKLQYENRVLLSNLQRCDIASCQSTRPMLETDAEAGDSAQCVPAPLGETHESHAV 420
 RLCRAREAEVLPGRLREQAALVSKAIDVLVADANGFTAGLRLCLDNECADPRLHEAPDNSE 480
 GPRDTKLIHArLVRLSVLQQELNAFTRKADAVLGCSVKEQOESFSSLPPGLSQGLSKEIL 540
 LAKDLGSDFQPDPFRDLPEWEPRIAREAFRTGDLDSKPDPSSRSPRPYRAEDNOSYASEIKE 600
 LQLVLAKAHDLSLRGLQEQLSQRQLRKKEADNFNQKHVQLKEDQQRALRREFELQSLGL 660

-continued

```

QRRLEQKFWSQEKNMLVQESQQFKHNFLLLFMKLRWFLKRWRQGKVLPEGDDFLEVNSM 720
KELYLLMEEEINAQHSDNKACTGDSWTQNTPNEYIKTLADMKVTLKELCWLLRDERGL 780
TELQQQFAKAKATWETERAELKGHTSQMELKTGKAGERAGPDWKAALQREREEQHLLA 840
KSYSAVMELTROLQISERNWSQKKLQLVERLQGEKQQVEQQVKELQNRLSQLQKAADPWW 900
LKHSELEKQDNWSKETRSEKIHDKEAVSEVELGGNGLKRTKSVSSMSEFESLLDCSPYIA 960
GGDARGKKLPNNPAFGFVSSEPGDPEKDTKEKPGLSSHUCNHLGALACDPPGRQMQRSY 1020
TAPDKTGIRVYYSPPVARRLGVPVVHDKEGKIIIEPGFLFTTAKPKESAKADGLAESSYG 1080
RWLCNFNSRQLDGGSAGPSAAGPGFPAAALHDFFEMSGNMSDDMKEITNCVRQAMRSGSLE 1140
RKVKSTSSQTVGGLASVGTQTIRTVSVGLQTDPPRSSLHGKAWSPRSSSLVSRSKQTSSS 1200
LDKVHSRIERPCCSPKyGSPKLQRRSVSKLDSSKDRSLWNLHQGKQNGSAWARSTTRDS 1250
PVLRNINDGLSSLFSVVEHSGSTESVWKLGMSKTRAKPEPPKYGIVQEFFFNRVCGRAPSP 1320
TSSAGEEGTKKPEPLSPASYHQPEGVARILNKAALKGSSEKVRLTMLPQVGKDGVIRDG 1380
DGAVVLPNEDAVCDCSTQSLTSCFARSSRSAIRHSPSKCRLHPSESSWGGEEERALPPSE 1439

```

[0075] One embodiment of the invention is an isolated polynucleotide selected from the group consisting of:

[0076] (a) a polynucleotide comprising a nucleotide sequence at least 70% (e.g., 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100%) identical to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3 and encoding a functional SOGA polypeptide;

[0077] (b) a polynucleotide that hybridizes to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3 under stringent hybridization conditions and encodes a functional SOGA polypeptide;

[0078] (c) a polynucleotide encoding a functional SOGA polypeptide comprising an amino acid sequence at least 70% (e.g., 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100%) identical to an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 4; and

[0079] (d) a functional fragment of any of (a) to (c).

[0080] In another embodiment, the isolated polynucleotide is selected from the group consisting of:

[0081] (a) a polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3 or a fragment thereof that encodes a functional SOGA polypeptide;

[0082] (b) a polynucleotide encoding a functional SOGA polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 4 or a functional fragment thereof; and

[0083] (c) a polynucleotide comprising a nucleotide sequence that differs from the nucleotide sequences of (a) or (b) above due to the degeneracy of the genetic code.

[0084] In one aspect, the invention relates to SOGA polypeptides and functional fragments or homologs thereof. The SOGA polypeptide can be from any species expressing SOGA, such as mammalian SOGA, e.g., human or mouse SOGA. As used herein, the term "homolog" is used to refer to a polypeptide which differs from a naturally occurring polypeptide by minor modifications to the naturally occurring polypeptide, but which significantly retains a biological activity of the naturally occurring polypeptide. Minor modifications include, without limitation, changes in one or a few

amino acid side chains, changes to one or a few amino acids (including deletions, insertions, and substitutions), changes in stereochemistry of one or a few atoms, and minor derivatizations, including, without limitation, methylation, glycosylation, phosphorylation, acetylation, myristoylation, prenylation, palmitation, amidation, and addition of glycosylphosphatidyl inositol. The term "substantially retains," as used herein, refers to a fragment, homolog, or other variant of a polypeptide that retains at least about 20% of the activity of the naturally occurring polypeptide (e.g., inhibition of glucose production), e.g., about 30%, 40%, 50% or more. SOGA activity can be measured as disclosed herein. Other biological activities may include enzyme activity, receptor binding, ligand binding, a cell signal transduction event, etc.

[0085] Functional fragments of SOGA polypeptide include any fragment that substantially retains at least one biological activity of full length SOGA polypeptide. In one embodiment, the functional fragment is a C-terminal fragment of SOGA. In certain embodiments, the C-terminal fragment begins immediately after the internal signal sequence of SOGA. In other embodiments, the functional fragment is a C-terminal fragment of about 80 kDa or 25 kDa.

[0086] In exemplary embodiments, the polypeptide comprises, consists essentially of, or consists of the amino acid sequence of the polypeptide disclosed herein and in the GenBank accession numbers listed above or a functional fragment thereof. In another embodiment, the isolated polypeptide comprises, consists essentially of, or consists of an amino acid sequence that is at least 70% identical, e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the disclosed amino acid sequence or a functional fragment thereof (and polynucleotide sequences encoding the same).

[0087] The polypeptide of the invention also include functional portions or fragments (and polynucleotide sequences encoding the same). The length of the fragment is not critical as long as it substantially retains at least one biological activity of the polypeptide. Illustrative fragments comprise at least

about 4, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, 300, 400, 500, or more contiguous amino acids of a SOGA polypeptide.

[0088] Likewise, those skilled in the art will appreciate that the present invention also encompasses fusion polypeptides (and polynucleotide sequences encoding the same) comprising a SOGA polypeptide or a functional fragment thereof. For example, it may be useful to express the polypeptide (or functional fragment) as a fusion protein that can be recognized by a commercially available antibody (e.g., FLAG motifs) or as a fusion protein that can otherwise be more easily purified (e.g., by addition of a poly-His tail). Additionally, fusion proteins that enhance the stability of the polypeptide may be produced, e.g., fusion proteins comprising maltose binding protein or glutathione-S-transferase. As another alternative, the fusion protein can comprise a reporter molecule. In other embodiments, the fusion protein can comprise a polypeptide that provides a function or activity that is the same as or different from the activity of the polypeptide, e.g., a targeting, binding, or enzymatic activity or function.

[0089] Likewise, it will be understood that the SOGA polypeptides specifically disclosed herein will typically tolerate substitutions in the amino acid sequence and substantially retain biological activity. To identify polypeptides of the invention other than those specifically disclosed herein, amino acid substitutions may be based on any characteristic known in the art, including the relative similarity or differences of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like.

[0090] Amino acid substitutions other than those disclosed herein may be achieved by changing the codons of the DNA sequence (or RNA sequence), according to the following codon table.

TABLE 1

Amino Acid		Codons
Alanine	Ala A	GCA GCC GCG GCT
Cysteine	Cys C	TGC TGT
Aspartic acid	Asp D	GAC GAT
Glutamic acid	Glu E	GAA GAG
Phenylalanine	Phe F	TTC TTT
Glycine	Gly G	GGA GGC GGG GGT
Histidine	His H	CAC CAT
Isoleucine	Ile I	ATA ATC ATT
Lysine	Lys K	AAA AAG
Leucine	Leu L	TTA TTG CTA CTC CTG CTT
Methionine	Met M	ATG
Asparagine	Asn N	AAC AAT
Proline	Pro P	CCA CCC CCG CCT
Glutamine	Gln Q	CAA CAG
Arginine	Arg R	AGA AGG CGA CGC CGG CGT
Serine	Ser S	AGC ACT TCA TCC TCG TCT

TABLE 1-continued

Amino Acid		Codons
Threonine	Thr T	ACA ACC ACG ACT
Valine	Val V	GTA GTC GTG GTT
Tryptophan	Trp W	TGG
Tyrosine	Tyr Y	TAC TAT

[0091] In identifying amino acid sequences encoding polypeptides other than those specifically disclosed herein, the hydrophobic index of amino acids may be considered. The importance of the hydrophobic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (see, Kyte and Doolittle, *J. Mol. Biol.* 157:105 (1982); incorporated herein by reference in its entirety). It is accepted that the relative hydrophobic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

[0092] Each amino acid has been assigned a hydrophobic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, id.), these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); praline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5). Accordingly, the hydrophobic index of the amino acid (or amino acid sequence) may be considered when modifying the polypeptides specifically disclosed herein.

[0093] It is also understood in the art that the substitution of amino acids can be made on the basis of hydrophilicity. U.S. Pat. No. 4,554,101 (incorporated herein by reference in its entirety) states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

[0094] As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0±1); glutamate (+3.0±1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5±1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). Thus, the hydrophilicity of the amino acid (or amino acid sequence) may be considered when identifying additional polypeptides beyond those specifically disclosed herein.

[0095] In embodiments of the invention, the polynucleotide encoding the SOGA polypeptide (or functional fragment) will hybridize to the nucleic acid sequences specifically disclosed herein or fragments thereof under standard conditions as known by those skilled in the art and encode a functional polypeptide or functional fragment thereof.

[0096] For example, hybridization of such sequences may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions (e.g., conditions represented by a wash stringency of 35-40% formamide

with 5×Denhardt's solution, 0.5% SDS and 1×SSPE at 37° C.; conditions represented by a wash stringency of 40-45% formamide with 5×Denhardt's solution, 0.5% SDS, and 1×SSPE at 42° C.; and conditions represented by a wash stringency of 50% formamide with 5×Denhardt's solution, 0.5% SDS and 1×SSPE at 42° C., respectively) to the polynucleotide sequences encoding the SOGA polypeptides or functional fragments thereof specifically disclosed herein. See, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual 2nd Ed. (Cold Spring Harbor, N.Y., 1989).

[0097] In other embodiments, polynucleotide sequences encoding the SOGA polypeptides have at least about 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or higher sequence identity with the nucleic acid sequences disclosed herein and in the GenBank accession numbers listed above or functional fragments thereof and encode a functional polypeptide or functional fragment thereof.

[0098] Further, it will be appreciated by those skilled in the art that there can be variability in the polynucleotides that encode the polypeptides (and fragments thereof) of the present invention due to the degeneracy of the genetic code. The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same polypeptide, is well known in the literature (See, e.g., Table 1).

[0099] Likewise, the polypeptides (and fragments thereof) of the invention include polypeptides that have at least about 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or higher amino acid sequence identity with the disclosed polypeptide sequences.

[0100] As is known in the art, a number of different programs can be used to identify whether a polynucleotide or polypeptide has sequence identity or similarity to a known sequence. Sequence identity or similarity may be determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.), the Best Fit sequence program described by Devereux et al., *Nucl. Acid Res.* 12:387 (1984), preferably using the default settings, or by inspection.

[0101] An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351 (1987); the method is similar to that described by Higgins & Sharp, *CABIOS* 5:151 (1989).

[0102] Another example of a useful algorithm is the BLAST algorithm, described in Altschul et al., *J. Mol. Biol.* 215:403 (1990) and Karlin et al., *Proc. Natl. Acad. Sci. USA* 90:5873 (1993). A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul et al., *Meth. Enzymol.*, 266:460 (1996); blastmstl/edu/blast/ README.html. WU-BLAST-2 uses several search parameters, which are preferably set to the default values. The parameters are dynamic values and are established by the program itself depending upon the composition of the par-

ticular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity.

[0103] An additional useful algorithm is gapped BLAST as reported by Altschul et al., *Nucleic Acids Res.* 25:3389 (1997).

[0104] A percentage amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored).

[0105] In a similar manner, percent nucleic acid sequence identity with respect to the coding sequence of the polypeptides disclosed herein is defined as the percentage of nucleotide residues in the candidate sequence that are identical with the nucleotides in the polynucleotide specifically disclosed herein.

[0106] The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer amino acids than the polypeptides specifically disclosed herein, it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical amino acids in relation to the total number of amino acids. Thus, for example, sequence identity of sequences shorter than a sequence specifically disclosed herein, will be determined using the number of amino acids in the shorter sequence, in one embodiment. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as insertions, deletions, substitutions, etc.

[0107] In one embodiment, only identities are scored positively (+1) and all forms of sequence variation including gaps are assigned a value of "0," which obviates the need for a weighted scale or parameters as described below for sequence similarity calculations. Percent sequence identity can be calculated, for example, by dividing the number of matching identical residues by the total number of residues of the "shorter" sequence in the aligned region and multiplying by 100. The "longer" sequence is the one having the most actual residues in the aligned region.

[0108] Those skilled in the art will appreciate that the isolated polynucleotides encoding the polypeptides of the invention will typically be associated with appropriate expression control sequences, e.g., transcription/translation control signals and polyadenylation signals.

[0109] It will further be appreciated that a variety of promoter/enhancer elements can be used depending on the level and tissue-specific expression desired. The promoter can be constitutive or inducible, depending on the pattern of expression desired. The promoter can be native or foreign and can be a natural or a synthetic sequence. By foreign, it is intended that the transcriptional initiation region is not found in the wild-type host into which the transcriptional initiation region is introduced. The promoter is chosen so that it will function in the target cell(s) of interest.

[0110] To illustrate, the polypeptide coding sequence can be operatively associated with a cytomegalovirus (CMV) major immediate-early promoter, an albumin promoter, an Elongation Factor 1- α (EF1- α) promoter, a PyK promoter, a MFG promoter, or a Rous sarcoma virus promoter.

[0111] Inducible promoter/enhancer elements include hormone-inducible and metal-inducible elements, and other pro-

motors regulated by exogenously supplied compounds, including without limitation, the zinc-inducible metallothionein (MT) promoter; the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter; the T7 polymerase promoter system (see WO 98/10088); the ecdysone insect promoter (No et al., *Proc. Natl. Acad. Sci. USA* 93:3346 (1996)); the tetracycline-repressible system (Gossen et al., *Proc. Natl. Acad. Sci. USA* 89:5547 (1992)); the tetracycline-inducible system (Gossen et al., *Science* 268:1766 (1995); see also Harvey et al., *Curr. Opin. Chem. Biol.* 2:512 (1998)); the RU486-inducible system (Wang et al., *Nat. Biotech.* 15:239 (1997); Wang et al., *Gene Ther.* 4:432 (1997)); and the rapamycin-inducible system (Magari et al., *J. Clin. Invest.* 100:2865 (1997)).

[0112] Other tissue-specific promoters or regulatory promoters include, but are not limited to, promoters that typically confer tissue-specificity in hepatocytes. These include, but are not limited to, promoters for albumin, hepatocyte nuclear factors, transthyretin, α_1 -antitrypsin, and the hepatitis B virus core promoter. In other embodiments, the promoters typically confer tissue specific in renal cells. These include, but are not limited to, promoters for ksp-cadherin, erythropoietin, γ -glutamyl transpeptidase, kidney androgen-regulated protein, vacuolar H⁺-ATPase B1 subunit, and AQP2. In other embodiments, the promoters typically confer tissue specific in muscle cells, e.g., skeletal muscle and/or cardiac muscle. Skeletal muscle cell promoters include, but are not limited to, promoters for β -actin, Pitx3, creatine kinase, and myosin light chain. Cardiac muscle cell promoters include, but are not limited to, promoters for cardiac actin, cardiac troponin T, troponin C, myosin light chain-2, and a-myosin heavy chain.

[0113] Moreover, specific initiation signals are generally required for efficient translation of inserted polypeptide coding sequences. These translational control sequences, which can include the ATG initiation codon and adjacent sequences, can be of a variety of origins, both natural and synthetic.

[0114] The present invention further provides cells comprising the isolated polynucleotides and polypeptides of the invention. The cell may be a cultured cell or a cell *in vivo*, e.g., for use in therapeutic methods, diagnostic methods, screening methods, methods for studying the biological action of SOGA polypeptides, in methods of producing the polypeptides, or in methods of maintaining or amplifying the polynucleotides of the invention, etc. In another embodiment, the cell is an *ex vivo* cell that has been isolated from a subject. The *ex vivo* cell may be modified and then reintroduced into the subject for diagnostic or therapeutic purposes.

[0115] In particular embodiments, the cell is an untransformed cell or a cell from a cell line of a gluconeogenic tissue, such as liver, kidney, skeletal muscle, or cardiac muscle.

[0116] The isolated polynucleotide can be incorporated into an expression vector. Expression vectors compatible with various host cells are well known in the art and contain suitable elements for transcription and translation of nucleic acids. Typically, an expression vector contains an "expression cassette," which includes, in the 5' to 3' direction, a promoter, a coding sequence encoding a SOGA polypeptide or functional fragment thereof operatively associated with the promoter, and, optionally, a termination sequence including a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase.

[0117] Non-limiting examples of promoters of this invention include CYC1, E1TS3, GAL1, GAL4, GAL10, ADH1, PGK, P1105, GAPDH, ADC1, TRP1, URA3, LEU2, ENO,

TPI, and alkaline phosphatase promoters (useful for expression in *Saccharomyces*); AOX1 promoter (useful for expression in *Pichia*); β -lactamase, lac, ara, tet, trp, IP_L, IP_R, T7, tac, and trc promoters (useful for expression in *Escherichia coli*); light regulated-, seed specific-, pollen specific-, ovary specific-, pathogenesis or disease related-promoters, cauliflower mosaic virus ³⁵S, CMV ³⁵S minimal, cassava vein mosaic virus (CsVMV), chlorophyll a/b binding protein, ribulose 1,5-bisphosphate carboxylase, shoot-specific promoters, root specific promoters, chitinase, stress inducible promoters, rice tungro bacilliform virus, plant super-promoter, potato leucine aminopeptidase, nitrate reductase, mannopine synthase, nopaline synthase, ubiquitin, zein protein, and anthocyanin promoters (useful for expression in plant cells).

[0118] Further examples of animal and mammalian promoters known in the art include, but are not limited to, the SV40 early (SV40e) promoter region, the promoter contained in the 3' long terminal repeat (LTR) of Rous sarcoma virus (RSV), the promoters of the E1A or major late promoter (MLP) genes of adenoviruses (Ad), the cytomegalovirus (CMV) early promoter, the herpes simplex virus (HSV) thymidine kinase (TK) promoter, baculovirus IE1 promoter, elongation factor 1 alpha (EF1) promoter, phosphoglycerate kinase (PGK) promoter, ubiquitin (Ubc) promoter, an albumin promoter, the regulatory sequences of the mouse metallothionein-L promoter and transcriptional control regions, the ubiquitous promoters (HPRT, vimentin, α -actin, tubulin and the like), the promoters of the intermediate filaments (desmin, neurofilaments, keratin, GFAP, and the like), the promoters of therapeutic genes (of the MDR, CFTR or factor VIII type, and the like), pathogenesis and/or disease-related promoters, and promoters that exhibit tissue specificity, such as the elastase I gene control region, which is active in pancreatic acinar cells; the insulin gene control region active in pancreatic beta cells, the immunoglobulin gene control region active in lymphoid cells, the mouse mammary tumor virus control region active in testicular, breast, lymphoid and mast cells; the albumin gene promoter, the Apo AI and Apo All control regions active in liver, the alpha-fetoprotein gene control region active in liver, the alpha 1-antitrypsin gene control region active in the liver, the beta-globin gene control region active in myeloid cells, the myelin basic protein gene control region active in oligodendrocyte cells in the brain, the myosin light chain-2 gene control region active in skeletal muscle, and the gonadotropic releasing hormone gene control region active in the hypothalamus, the pyruvate kinase promoter, the villin promoter, the promoter of the fatty acid binding intestinal protein, the promoter of smooth muscle cell α -actin, and the like. In addition, any of these expression sequences of this invention can be modified by addition of enhancer and/or regulatory sequences and the like.

[0119] Enhancers that may be used in embodiments of the invention include but are not limited to: an SV40 enhancer, a cytomegalovirus (CMV) enhancer, an elongation factor 1 (EF1) enhancer, yeast enhancers, viral gene enhancers, and the like.

[0120] Termination control regions, i.e., terminator or polyadenylation sequences, may be derived from various genes native to the preferred hosts. In some embodiments of the invention, the termination control region may comprise or be derived from a synthetic sequence, a synthetic polyadenylation signal, an SV40 late polyadenylation signal, an SV40

polyadenylation signal, a bovine growth hormone (BGH) polyadenylation signal, viral terminator sequences, or the like.

[0121] It will be apparent to those skilled in the art that any suitable vector can be used to deliver the polynucleotide to a cell or subject. The vector can be delivered to cells *in vivo*. In other embodiments, the vector can be delivered to cells *ex vivo*, and then cells containing the vector are delivered to the subject. The choice of delivery vector can be made based on a number of factors known in the art, including age and species of the target host, *in vitro* versus *in vivo* delivery, level and persistence of expression desired, intended purpose (e.g., for therapy or screening), the target cell or organ, route of delivery, size of the isolated polynucleotide, safety concerns, and the like.

[0122] Suitable vectors include plasmid vectors, viral vectors (e.g., retrovirus, alphavirus; vaccinia virus; adenovirus, adeno-associated virus and other parvoviruses, lentivirus, poxvirus, or herpes simplex virus), lipid vectors, poly-lysine vectors, synthetic polyamino polymer vectors, and the like.

[0123] Any viral vector that is known in the art can be used in the present invention. Protocols for producing recombinant viral vectors and for using viral vectors for nucleic acid delivery can be found in Ausubel et al., *Current Protocols in Molecular Biology* (Green Publishing Associates, Inc. and John Wiley & Sons, Inc., New York) and other standard laboratory manuals (e.g., *Vectors for Gene Therapy*. In: *Current Protocols in Human Genetics*. John Wiley and Sons, Inc.: 1997).

[0124] Non-viral transfer methods can also be employed. Many non-viral methods of nucleic acid transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In particular embodiments, non-viral nucleic acid delivery systems rely on endocytic pathways for the uptake of the nucleic acid molecule by the targeted cell. Exemplary nucleic acid delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

[0125] In particular embodiments, plasmid vectors are used in the practice of the present invention. For example, naked plasmids can be introduced into muscle cells by injection into the tissue. Expression can extend over many months, although the number of positive cells is typically low (Wolff et al., *Science* 247:247 (1989)). Cationic lipids have been demonstrated to aid in introduction of nucleic acids into some cells in culture (Feigner and Ringold, *Nature* 337:387 (1989)). Injection of cationic lipid plasmid DNA complexes into the circulation of mice has been shown to result in expression of the DNA in lung (Brigham et al., *Am. J. Med. Sci.* 298:278 (1989)). One advantage of plasmid DNA is that it can be introduced into non-replicating cells.

[0126] In a representative embodiment, a nucleic acid molecule (e.g., a plasmid) can be entrapped in a lipid particle bearing positive charges on its surface and, optionally, tagged with antibodies against cell surface antigens of the target tissue (Mizuno et al., *No Shinkei Geka* 20:547 (1992); PCT publication WO 91/06309; Japanese patent application 1047381; and European patent publication EP-A-43075).

[0127] Liposomes that consist of amphiphilic cationic molecules are useful as non-viral vectors for nucleic acid delivery *in vitro* and *in vivo* (reviewed in Crystal, *Science* 270:404 (1995); Blaese et al., *Cancer Gene Ther.* 2:291 (1995); Behr et al., *Bioconjugate Chem.* 5:382 (1994); Remy et al., *Bioconjugate Chem.* 5:647 (1994); and Gao et al., *Gene Therapy*

2:710 (1995)). The positively charged liposomes are believed to complex with negatively charged nucleic acids via electrostatic interactions to form lipid:nucleic acid complexes. The lipid:nucleic acid complexes have several advantages as nucleic acid transfer vectors. Unlike viral vectors, the lipid:nucleic acid complexes can be used to transfer expression cassettes of essentially unlimited size. Since the complexes lack proteins, they can evoke fewer immunogenic and inflammatory responses. Moreover, they cannot replicate or recombine to form an infectious agent and have low integration frequency. A number of publications have demonstrated that amphiphilic cationic lipids can mediate nucleic acid delivery *in vivo* and *in vitro* (Feigner et al., *Proc. Natl. Acad. Sci. USA* 84:7413 (1987); Loeffler et al., *Meth. Enzymol.* 217:599 (1993); Feigner et al., *J. Biol. Chem.* 269:2550 (1994)).

[0128] Several groups have reported the use of amphiphilic cationic lipid:nucleic acid complexes for *in vivo* transfection both in animals and in humans (reviewed in Gao et al., *Gene Therapy* 2:710 (1995); Zhu et al., *Science* 261:209 (1993); and Thierry et al., *Proc. Natl. Acad. Sci. USA* 92:9742 (1995)). U.S. Pat. No. 6,410,049 describes a method of preparing cationic lipid:nucleic acid complexes that have a prolonged shelf life.

[0129] Expression vectors can be designed for expression of polypeptides in prokaryotic or eukaryotic cells. For example, polypeptides can be expressed in bacterial cells such as *E. coli*, insect cells (e.g., the baculovirus expression system), yeast cells, plant cells or mammalian cells. Some suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990). Examples of bacterial vectors include pQE70, pQE60, pQE-9 (Qiagen), pBS, pD10, phagescript, psiX174, pbluescript SK, pbks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, and pRIT5 (Pharmacia). Examples of vectors for expression in the yeast *S. cerevisiae* include pYEpSec1 (Baldari et al., *EMBO J.* 6:229 (1987)), pMFa (Kurjan and Herskowitz, *Cell* 30:933 (1982)), pJRY88 (Schultz et al., *Gene* 54:113 (1987)), and pYES2 (Invitrogen Corporation, San Diego, Calif.). Baculovirus vectors available for expression of nucleic acids to produce proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., *Mol. Cell. Biol.* 3:2156 (1983)) and the pVL series (Lucklow and Summers *Virology* 170:31 (1989)).

[0130] Examples of mammalian expression vectors include pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, PBPV, pMSG, PSVL (Pharmacia), pCDM8 (Seed, *Nature* 329:840 (1987)) and pMT2PC (Kaufman et al., *EMBO J.* 6:187 (1987)). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus and Simian Virus 40.

[0131] Viral vectors have been used in a wide variety of gene delivery applications in cells, as well as living animal subjects. Viral vectors that can be used include, but are not limited to, retrovirus, lentivirus, adeno-associated virus, poxvirus, alphavirus, baculovirus, vaccinia virus, herpes virus, Epstein-Barr virus, adenovirus, geminivirus, and caulimovirus vectors. Non-viral vectors include plasmids, liposomes, electrically charged lipids (cytoselectins), nucleic acid-protein complexes, and biopolymers. In addition to a nucleic acid of interest, a vector may also comprise one or more regulatory regions, and/or selectable markers useful in selecting, mea-

suring, and monitoring nucleic acid transfer results (delivery to specific tissues, duration of expression, etc.).

[0132] In addition to the regulatory control sequences discussed above, the recombinant expression vector can contain additional nucleotide sequences. For example, the recombinant expression vector can encode a selectable marker gene to identify host cells that have incorporated the vector.

[0133] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" refer to a variety of art-recognized techniques for introducing foreign nucleic acids (e.g., DNA and RNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, microinjection, DNA-loaded liposomes, lipofectamine-DNA complexes, cell sonication, gene bombardment using high velocity micro-projectiles, and viral-mediated transfection. Suitable methods for transforming or transfecting host cells can be found in Sambrook et al., Molecular Cloning: A Laboratory Manual 2nd Ed. (Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

[0134] If stable integration is desired, often only a small fraction of cells (in particular, mammalian cells) integrate the foreign DNA into their genome. In order to identify and select integrants, a nucleic acid that encodes a selectable marker (e.g., resistance to antibiotics) can be introduced into the host cells along with the nucleic acid of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acids encoding a selectable marker can be introduced into a host cell on the same vector as that comprising the nucleic acid of interest or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0135] Polypeptides and fragments of the invention can be modified for in vivo use by the addition, at the amino- and/or carboxyl-terminal ends, of a blocking agent to facilitate survival of the relevant polypeptide in vivo. This can be useful in those situations in which the peptide termini tend to be degraded by proteases prior to cellular uptake. Such blocking agents can include, without limitation, additional related or unrelated peptide sequences that can be attached to the amino and/or carboxyl terminal residues of the peptide to be administered. This can be done either chemically during the synthesis of the peptide or by recombinant DNA technology by methods familiar to artisans of average skill. Alternatively, blocking agents such as pyroglutamic acid or other molecules known in the art can be attached to the amino and/or carboxyl terminal residues, or the amino group at the amino terminus or carboxyl group at the carboxyl terminus can be replaced with a different moiety. Likewise, the peptides can be covalently or noncovalently coupled to pharmaceutically acceptable "carrier" proteins prior to administration.

[0136] Another embodiment of the invention relates to homologs of the polypeptides of the invention that are peptidomimetic compounds that are designed based upon the amino acid sequences of the functional polypeptide fragments. Peptidomimetic compounds are synthetic compounds having a three-dimensional conformation (i.e., a "peptide motif") that is substantially the same as the three-dimensional conformation of a selected peptide. The peptide motif provides the peptidomimetic compound with biological activi-

ties qualitatively identical to that of the functional fragment from which the peptidomimetic was derived. Peptidomimetic compounds can have additional characteristics that enhance their therapeutic utility, such as increased cell permeability and prolonged biological half-life.

[0137] The peptidomimetics typically have a backbone that is partially or completely non-peptide, but with side groups that are identical to the side groups of the amino acid residues that occur in the peptide on which the peptidomimetic is based. Several types of chemical bonds, e.g., ester, thioester, thioamide, retroamide, reduced carbon A, dimethylene and ketomethylene bonds, are known in the art to be generally useful substitutes for peptide bonds in the construction of protease-resistant peptidomimetics.

III. Inhibitors of SOGA Polypeptides and Polynucleotides

[0138] As one aspect, the invention provides agents that inhibit the expression and/or activity of SOGA polypeptides or polynucleotides. These agents can be used to inhibit or down-regulate the SOGA signaling pathway, e.g., in a cell or a subject.

[0139] In one embodiment of the invention, decreasing the expression and/or activity of a SOGA polypeptide comprises decreasing the level of a nucleic acid (DNA or RNA) encoding the polypeptide or the level of expression of the polypeptide from the nucleic acid. Numerous methods for reducing the level and/or expression of polynucleotides in vitro or in vivo are known. For example, the nucleotide sequences for the human and mouse SOGA polypeptides are disclosed herein. An antisense nucleotide sequence or nucleic acid encoding an antisense nucleotide sequence can be generated to any portion thereof in accordance with known techniques.

[0140] The term "antisense nucleotide sequence" or "antisense oligonucleotide" as used herein, refers to a nucleotide sequence that is complementary to a specified DNA or RNA sequence. Antisense oligonucleotides and nucleic acids that express the same can be made in accordance with conventional techniques. See, e.g., U.S. Pat. No. 5,023,243 to Tullis; U.S. Pat. No. 5,149,797 to Pederson et al. The antisense nucleotide sequence can be complementary to the entire nucleotide sequence encoding the polypeptide or a portion thereof of at least 10, 20, 40, 50, 75, 100, 150, 200, 300, or 500 contiguous bases or more and will reduce the level of polypeptide production.

[0141] Those skilled in the art will appreciate that it is not necessary that the antisense nucleotide sequence be fully complementary to the target sequence as long as the degree of sequence similarity is sufficient for the antisense nucleotide sequence to hybridize to its target and reduce production of the polypeptide. As is known in the art, a higher degree of sequence similarity is generally required for short antisense nucleotide sequences, whereas a greater degree of mismatched bases will be tolerated by longer antisense nucleotide sequences.

[0142] For example, hybridization of such nucleotide sequences can be carried out under conditions of reduced stringency, medium stringency or even stringent conditions (e.g., conditions represented by a wash stringency of 35-40% formamide with 5×Denhardt's solution, 0.5% SDS and 1×SSPE at 37° C.; conditions represented by a wash stringency of 40-45% formamide with 5×Denhardt's solution, 0.5% SDS, and 1×SSPE at 42° C.; and/or conditions represented by a wash stringency of 50% formamide with 5×Den-

hardt's solution, 0.5% SDS and 1×SSPE at 42° C., respectively) to the nucleotide sequences specifically disclosed herein. See, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual 2nd Ed. (Cold Spring Harbor, N.Y., 1989).

[0143] In other embodiments, antisense nucleotide sequences of the invention have at least about 70%, 80%, 90%, 95%, 97%, 98% or higher sequence similarity with the complement of the coding sequences specifically disclosed herein and will reduce the level of polypeptide production.

[0144] In other embodiments, the antisense nucleotide sequence can be directed against any coding sequence, the silencing of which results in a modulation of a SOGA polypeptide.

[0145] The length of the antisense nucleotide sequence (i.e., the number of nucleotides therein) is not critical as long as it binds selectively to the intended location and reduces transcription and/or translation of the target sequence, and can be determined in accordance with routine procedures. In general, the antisense nucleotide sequence will be from about eight, ten or twelve nucleotides in length up to about 20, 30, 50, 75 or 100 nucleotides, or longer, in length.

[0146] An antisense nucleotide sequence can be constructed using chemical synthesis and enzymatic ligation reactions by procedures known in the art. For example, an antisense nucleotide sequence can be chemically synthesized using naturally occurring nucleotides or various modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleotide sequences, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleotide sequence include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyl-2-thiouracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethyl guanine, 2-methylenadenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiacytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N2-carboxypropyl)uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleotide sequence can be produced using an expression vector into which a nucleic acid has been cloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest).

[0147] The antisense nucleotide sequences of the invention further include nucleotide sequences wherein at least one, or all, of the internucleotide bridging phosphate residues are modified phosphates, such as methyl phosphonates, methyl phosphonothioates, phosphoromorpholidates, phosphoropiperazidates and phosphoramidates. For example, every other one of the internucleotide bridging phosphate residues can be modified as described. In another non-limiting example, the antisense nucleotide sequence is a nucleotide

sequence in which one, or all, of the nucleotides contain a 2' lower alkyl moiety (e.g., C₁-C₄, linear or branched, saturated or unsaturated alkyl, such as methyl, ethyl, ethenyl, propyl, 1-propenyl, 2-propenyl, and isopropyl). For example, every other one of the nucleotides can be modified as described. See also, Furdon et al., *Nucleic Acids Res.* 17:9193 (1989); Agrawal et al., *Proc. Natl. Acad. Sci. USA* 87:1401 (1990); Baker et al., *Nucleic Acids Res.* 18:3537 (1990); Sproat et al., *Nucleic Acids Res.* 17:3373 (1989); Walder and Walder, *Proc. Natl. Acad. Sci. USA* 85:5011 (1988); incorporated by reference herein in their entireties for their teaching of methods of making antisense molecules, including those containing modified nucleotide bases).

[0148] Triple helix base-pairing methods can also be employed to inhibit production of SOGA polypeptides. Triple helix pairing is believed to work by inhibiting the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (e.g., Gee et al., (1994) In: Huber et al., Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.).

[0149] Small Interference (si) RNA, also known as RNA interference (RNAi) molecules, provides another approach for modulating the expression of SOGA polypeptides. The siRNA can be directed against polynucleotide sequences encoding the SOGA polypeptides or any other sequence that results in modulation of the expression of SOGA polypeptides.

[0150] siRNA is a mechanism of post-transcriptional gene silencing in which double-stranded RNA (dsRNA) corresponding to a coding sequence of interest is introduced into a cell or an organism, resulting in degradation of the corresponding mRNA. The mechanism by which siRNA achieves gene silencing has been reviewed in Sharp et al., *Genes Dev.* 15:485 (2001); and Hammond et al., *Nature Rev. Gen.* 2:110 (2001)). The siRNA effect persists for multiple cell divisions before gene expression is regained. siRNA is therefore a powerful method for making targeted knockouts or "knock-downs" at the RNA level. siRNA has proven successful in human cells, including human embryonic kidney and HeLa cells (see, e.g., Elbashir et al., *Nature* 411:494 (2001)). In one embodiment, silencing can be induced in mammalian cells by enforcing endogenous expression of RNA hairpins (see Padisson et al., *Proc. Natl. Acad. Sci. USA* 99:1443 (2002)). In another embodiment, transfection of small (21-23 nt) dsRNA specifically inhibits nucleic acid expression (reviewed in Caplen, *Trends Biotechnol.* 20:49 (2002)).

[0151] siRNA technology utilizes standard molecular biology methods. dsRNA corresponding to all or a part of a target coding sequence to be inactivated can be produced by standard methods, e.g., by simultaneous transcription of both strands of a template DNA (corresponding to the target sequence) with T7 RNA polymerase. Kits for production of dsRNA for use in siRNA are available commercially, e.g., from New England Biolabs, Inc. Methods of transfection of dsRNA or plasmids engineered to make dsRNA are routine in the art.

[0152] MicroRNA (miRNA), single stranded RNA molecules of about 21-23 nucleotides in length, can be used in a similar fashion to siRNA to modulate gene expression (see U.S. Pat. No. 7,217,807).

[0153] Silencing effects similar to those produced by siRNA have been reported in mammalian cells with transfec-

tion of a mRNA-cDNA hybrid construct (Lin et al., *Biochem. Biophys. Res. Commun.* 281:639 (2001)), providing yet another strategy for silencing a coding sequence of interest.

[0154] The expression of SOGA polypeptides can also be inhibited using ribozymes. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim et al., *Proc. Natl. Acad. Sci. USA* 84:8788 (1987); Gerlach et al., *Nature* 328:802 (1987); Forster and Symons, *Cell* 49:211 (1987)). For example, a large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Michel and Westhof, *J. Mol. Biol.* 216:585 (1990); Reinhold-Ilurek and Shub, *Nature* 357:173 (1992)). This specificity has been attributed to the requirement that the substrate bind via specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

[0155] Ribozyme catalysis has primarily been observed as part of sequence-specific cleavage/ligation reactions involving nucleic acids (Joyce, *Nature* 338:217 (1989)). For example, U.S. Pat. No. 5,354,855 reports that certain ribozymes can act as endonucleases with a sequence specificity greater than that of known ribonucleases and approaching that of the DNA restriction enzymes. Thus, sequence-specific ribozyme-mediated inhibition of gene expression may be particularly suited to therapeutic applications (Scanlon et al., *Proc. Natl. Acad. Sci. USA* 88:10591 (1991); Sarver et al., *Science* 247:1222 (1990); Sioud et al., *J. Mol. Biol.* 223:831 (1992)).

[0156] In another embodiment of the invention, decreasing the expression and/or activity of SOGA polypeptides comprises decreasing the activity of the polypeptide. Polypeptide activity can be modulated by interaction with an antibody or antibody fragment. The antibody or antibody fragment can bind to the polypeptide or to any other polypeptide of interest, as long as the binding between the antibody or the antibody fragment and the target polypeptide results in modulation of the activity of the SOGA polypeptide.

[0157] The term "antibody" or "antibodies" as used herein refers to all types of immunoglobulins, including IgG, IgM, IgA, IgD, and IgE. The antibody can be monoclonal or polyclonal and can be of any species of origin, including (for example) mouse, rat, rabbit, horse, goat, sheep, camel, or human, or can be a chimeric antibody. See, e.g., Walker et al., *Molec. Immunol.* 26:403 (1989). The antibodies can be recombinant monoclonal antibodies produced according to the methods disclosed in U.S. Pat. No. 4,474,893 or U.S. Pat. No. 4,816,567. The antibodies can also be chemically constructed according to the method disclosed in U.S. Pat. No. 4,676,980.

[0158] Antibody fragments included within the scope of the present invention include, for example, Fab, Fab', F(ab')₂, and Fv fragments; domain antibodies, diabodies; vaccibodies, linear antibodies; single-chain antibody molecules; and multispecific antibodies formed from antibody fragments. Such fragments can be produced by known techniques. For example, F(ab')₂ fragments can be produced by pepsin digestion of the antibody molecule, and Fab fragments can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries can be con-

structed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse et al., *Science* 254:1275 (1989)).

[0159] Antibodies of the invention may be altered or mutated for compatibility with species other than the species in which the antibody was produced. For example, antibodies may be humanized or camelized. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementarity determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework (FR) regions (i.e., the sequences between the CDR regions) are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., *Nature* 321:522 (1986); Riechmann et al., *Nature*, 332:323 (1988); and Presta, *Curr. Op. Struct. Biol.* 2:593 (1992)).

[0160] Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can essentially be performed following the method of Winter and co-workers (Jones et al., *Nature* 321:522 (1986); Riechmann et al., *Nature* 332:323 (1988); Verhoeyen et al., *Science* 239:1534 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Pat. No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues (e.g., all of the CDRs or a portion thereof) and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

[0161] Human antibodies can also be produced using various techniques known in the art, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.* 227:381 (1991); Marks et al., *J. Mol. Biol.* 222:581 (1991)). The techniques of Cole et al, and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, p. 77 (1985) and Boerner et al., *J. Immunol.* 147:86 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have

been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Pat. Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific publications: Marks et al., *Bio/Technology* 10:779 (1992); Lonberg et al., *Nature* 368: 856 (1994); Morrison, *Nature* 368:812 (1994); Fishwild et al., *Nature Biotechnol.* 14:845 (1996); Neuberger, *Nature Biotechnol.* 14:826 (1996); Lonberg and Huszar, *Intern. Rev. Immunol.* 13:65 (1995).

[0162] Polyclonal antibodies used to carry out the present invention can be produced by immunizing a suitable animal (e.g., rabbit, goat, etc.) with an antigen to which a monoclonal antibody to the target binds, collecting immune serum from the animal, and separating the polyclonal antibodies from the immune serum, in accordance with known procedures.

[0163] Monoclonal antibodies used to carry out the present invention can be produced in a hybridoma cell line according to the technique of Kohler and Milstein, *Nature* 265:495 (1975). For example, a solution containing the appropriate antigen can be injected into a mouse and, after a sufficient time, the mouse sacrificed and spleen cells obtained. The spleen cells are then immortalized by fusing them with myeloma cells or with lymphoma cells, typically in the presence of polyethylene glycol, to produce hybridoma cells. The hybridoma cells are then grown in a suitable medium and the supernatant screened for monoclonal antibodies having the desired specificity. Monoclonal Fab fragments can be produced in *E. coli* by recombinant techniques known to those skilled in the art. See, e.g., Huse, *Science* 246:1275 (1989).

[0164] Antibodies specific to the target polypeptide can also be obtained by phage display techniques known in the art.

[0165] Various immunoassays can be used for screening to identify antibodies having the desired specificity for the polypeptides of this invention. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificity are well known in the art. Such immunoassays typically involve the measurement of complex formation between an antigen and its specific antibody (e.g., antigen/antibody complex formation). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on the polypeptides or peptides of this invention can be used as well as a competitive binding assay.

[0166] Antibodies can be conjugated to a solid support (e.g., beads, plates, slides or wells formed from materials such as latex or polystyrene) in accordance with known techniques. Antibodies can likewise be conjugated to detectable groups such as radiolabels (e.g., ^{35}S , ^{125}I , ^{131}I), enzyme labels (e.g., horseradish peroxidase, alkaline phosphatase), and fluorescence labels (e.g., fluorescein) in accordance with known techniques. Determination of the formation of an antibody/antigen complex in the methods of this invention can be by detection of, for example, precipitation, agglutination, flocculation, radioactivity, color development or change, fluorescence, luminescence, etc., as is well known in the art.

[0167] In one embodiment, the activity of SAGA polypeptides is inhibited using aptamers. Recently, small structured single-stranded RNAs, also known as RNA aptamers, have emerged as viable alternatives to small-molecule and antibody-based therapy (Que-Gewirth et al., *Gene Ther.* 14:283

(2007); Ireson et al., *Mol. Cancer Ther.* 5:2957 (2006)). RNA aptamers specifically bind target proteins with high affinity, are quite stable, lack immunogenicity, and elicit biological responses. Aptamers are evolved by means of an iterative selection method called SELEX (systematic evolution of ligands by exponential enrichment) to specifically recognize and tightly bind their targets by means of well-defined complementary three-dimensional structures.

[0168] RNA aptamers represent a unique emerging class of therapeutic agents (Que-Gewirth et al., *Gene Ther.* 14:283 (2007); Ireson et al., *Mol. Cancer Ther.* 5:2957 (2006)). They are relatively short (12-30 nucleotide) single-stranded RNA oligonucleotides that assume a stable three-dimensional shape to tightly and specifically bind selected protein targets to elicit a biological response. In contrast to antisense oligonucleotides, RNA aptamers can effectively target extracellular targets. Like antibodies, aptamers possess binding affinities in the low nanomolar to picomolar range. In addition, aptamers are heat stable, lack immunogenicity, and possess minimal interbatch variability. Chemical modifications, such as amino or fluoro substitutions at the 2' position of pyrimidines, may reduce degradation by nucleases. The biodistribution and clearance of aptamers can also be altered by chemical addition of moieties such as polyethylene glycol and cholesterol. Further, SELEX allows selection from libraries consisting of up to 10^{15} ligands to generate high-affinity oligonucleotide ligands to purified biochemical targets.

[0169] In another embodiment, the method of decreasing the activity of a SOGA polypeptide comprises delivering to a cell or to a subject an agent that decreases the activity of a SOGA polypeptide, the agent administered in an amount effective to modulate the activity of the polypeptide. The agent can interact directly with the SOGA polypeptide to decrease the activity of the polypeptide. Alternatively, the agent can interact with any other polypeptide, nucleic acid or other molecule if such interaction results in a decrease of the activity of the SOGA.

[0170] The term "agent" as used herein is intended to be interpreted broadly and encompasses organic and inorganic molecules. Organic compounds include, but are not limited to, small molecules, polypeptides, lipids, carbohydrates, coenzymes, aptamers, and nucleic acid molecules (e.g., gene delivery vectors, antisense oligonucleotides, siRNA, all as described above).

[0171] Polypeptides include, but are not limited to, antibodies (described in more detail above) and enzymes. Nucleic acids include, but are not limited to, DNA, RNA and DNA-RNA chimeric molecules. Suitable RNA molecules include siRNA, antisense RNA molecules and ribozymes (all of which are described in more detail above). The nucleic acid can further encode any polypeptide such that administration of the nucleic acid and production of the polypeptide results in a decrease of the activity of a SOGA polypeptide.

[0172] The agent can further be an agent that is identified by any of the screening methods described below.

[0173] In one embodiment of the invention, the agent is a modulator of the insulin and/or adiponectin signaling pathways that directly or indirectly inhibits SOGA expression and/or activity. For example, the agent can be an activator of AMPK such as AICAR (N1-(β -D-ribofuranosyl)-5-aminoimidazole-4-carboxamide). In another embodiment, the

agent can be a P13 kinase inhibitor such as LY294002. In a further embodiment, the agent can be an inhibitor of adiponectin such as rapamycin.

IV. Inhibition of Glucose Production

[0174] Increases in SOGA polypeptide levels and/or activity result in the inhibition of glucose production in cells. Thus, the SOGA polypeptides and polynucleotides of the invention can be used in methods in which a decrease in glucose production is desired for research, diagnostic, and/or therapeutic proposes. These methods can be carried using techniques to increase the expression and/or activity of SOGA polypeptides in a cell, in a tissue, and/or in a subject.

[0175] One aspect of the invention relates to a method of decreasing glucose production in a cell, comprising contacting said cell with a polynucleotide, polypeptide, or fusion protein of the invention in an amount effective to decrease glucose production in the cell.

[0176] Another aspect of the invention relates to a method of decreasing autophagy in a cell, comprising contacting said cell with a polynucleotide, polypeptide, or fusion protein of the invention in an amount effective to decrease autophagy in said cell.

[0177] The cells to be contacted can be in vitro, ex vivo, or in vivo (e.g., in an animal model of disease or a patient). Cells can be contacted with a polynucleotide or polypeptide of the invention by any means known in the art and as described herein.

[0178] A further aspect of the invention relates to a method of decreasing blood glucose levels in a subject, comprising delivering to said subject a polynucleotide, polypeptide, or fusion protein of the invention in an amount effective to decrease the blood glucose levels in said subject.

[0179] Another aspect of the invention relates to a method of increasing insulin sensitivity in a subject, comprising delivering to said subject a polynucleotide, polypeptide, or fusion protein of the invention in an amount effective to increase insulin sensitivity in said subject.

[0180] In one embodiment, the subject is one that is in need of decreased glucose levels and/or increased insulin sensitivity. The subject can currently have or be at risk for a carbohydrate-related metabolic disorder such as diabetes mellitus (Type I or Type II), alcoholic ketoacidosis, diabetic ketoacidosis, nonketotic hyperosmolar syndrome, and new onset diabetes (NOD), such as in cancer patients undergoing chemotherapy, immunosuppressed patients, post-operative patients, and trauma patients. In certain embodiments, the methods of the invention encompass methods of treating a subject having a carbohydrate-related metabolic disorder such as diabetes, comprising delivering to said subject a polynucleotide, polypeptide, or fusion protein of the invention in an amount effective to treat the disorder.

[0181] In one embodiment, increasing the expression and/or activity of a SOGA polypeptide comprises delivering a nucleic acid encoding the polypeptide or a fragment or homolog thereof to the cell or tissue or subject. In another embodiment, increasing the expression and/or activity of a SOGA polypeptide comprises delivering the polypeptide itself or a fragment or homolog thereof to the cell or tissue or subject.

[0182] In one embodiment, the methods comprise delivering to the subject an isolated SOGA polypeptide. In exemplary embodiments, the polypeptide comprises, consists essentially of or consists of the amino acid sequence of the

polypeptide disclosed herein or a functional fragment thereof. In another embodiment, the isolated polypeptide comprises, consists essentially of, or consists of an amino acid sequence that is at least 70% identical, e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the disclosed amino acid sequence or a functional fragment thereof (and polynucleotide sequences encoding the same).

[0183] In one embodiment, the polynucleotides, polypeptides, or homologs thereof of the invention are administered directly to the subject. Generally, the compounds of the invention will be suspended in a pharmaceutically-acceptable carrier (e.g., physiological saline) and administered orally or by intravenous infusion, or injected subcutaneously, intramuscularly, intrathecally, intraperitoneally, intrarectally, intravaginally, intranasally, intragastrically, intratracheally, or intrapulmonarily. They can be delivered directly to a site involved in gluconeogenesis, such as the liver, kidney, and/or muscle. The dosage required depends on the choice of the route of administration; the nature of the formulation; the nature of the patient's illness; the subject's size, weight, surface area, age, and sex; other drugs being administered; and the judgment of the attending physician. Suitable dosages are in the range of 0.01-100.0 µg/kg. Wide variations in the needed dosage are to be expected in view of the variety of polypeptides and fragments available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by i.v. injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization as is well understood in the art. Administrations can be single or multiple (e.g., 2-, 3-, 4-, 6-, 8-, 10-, 20-, 50-, 100-, 150-, or more fold). Encapsulation of the polypeptide in a suitable delivery vehicle (e.g., polymeric microparticles or implantable devices) may increase the efficiency of delivery, particularly for oral delivery.

[0184] According to certain embodiments, the polynucleotides or vectors can be targeted to specific cells or tissues in vivo. Targeting delivery vehicles, including liposomes and viral vector systems are known in the art. For example, a liposome can be directed to a particular target cell or tissue by using a targeting agent, such as an antibody, soluble receptor or ligand, incorporated with the liposome, to target a particular cell or tissue to which the targeting molecule can bind. Targeting liposomes are described, for example, in Ho et al., *Biochemistry* 25:5500 (1986); Ho et al., *J. Biol. Chem.* 262: 13979 (1987); Ho et al., *J. Biol. Chem.* 262:13973 (1987); and U.S. Pat. No. 4,957,735 to Huang et al., each of which is incorporated herein by reference in its entirety). Enveloped viral vectors can be modified to deliver a nucleic acid molecule to a target cell by modifying or substituting an envelope protein such that the virus infects a specific cell type. In adenoviral vectors, the gene encoding the attachment fibers can be modified to encode a protein domain, that binds to a cell-specific receptor. Herpesvirus vectors naturally target the cells of the central and peripheral nervous system. Alternatively, the route of administration can be used to target a specific cell or tissue. For example, intracoronary administration of an adenoviral vector has been shown to be effective for the delivery of a gene to cardiac myocytes (Maurice et al., *J. Clin. Invest.* 104:21 (1999)). Intravenous delivery of cholesterol-containing cationic liposomes has been shown to preferentially target pulmonary tissues (Liu et al., *Nature Biotechnol.* 15:167 (1997)), and effectively mediate transfer and expression of genes in vivo. Other examples of successful

targeted in vivo delivery of nucleic acid molecules are known in the art. Finally, a recombinant nucleic acid molecule can be selectively (i.e., preferentially, substantially exclusively) expressed in a target cell by selecting a transcription control sequence, and preferably, a promoter, which is selectively induced in the target cell and remains substantially inactive in non-target cells.

[0185] The polypeptides and polynucleotides of the present invention can optionally be delivered in conjunction with other therapeutic agents. The additional therapeutic agents can be delivered concurrently with the polypeptides and polynucleotides of the invention. As used herein, the word "concurrently" means sufficiently close in time to produce a combined effect (that is, concurrently can be simultaneously, or it can be two or more events occurring within a short time period before or after each other). In one embodiment, the polypeptides and polynucleotides of the invention are administered in conjunction with anti-diabetic agents, including without limitation, (1) PPAR γ agonists such as glitazones (e.g., ciglitazone, darglitazone, englitazone, isaglitazone (MCC-555), pioglitazone, rosiglitazone, troglitazone, BRL49653, CLX-0921, 5-BTZD, GW-0207, LG-100641, and LY-300512; (2) biguanides such as buformin, metformin, and phenformin; (3) protein tyrosine phosphatase-1B (PTP-1B) inhibitors such as ISIS 113715; (4) sulfonylureas such as acetohexamide, chlorpropamide, diabinese, glibenclamide, glypizide, glyburide, glimepiride, gliclazide, glipentide, gliquidone, glisolamide, tolazamide, and tolbutamide; (5) meglitinides such as repaglinide and nateglinide; (6) alpha glucosidase hydrolase inhibitors such as acarbose, adiposine, camiglibose, emiglitate, miglitol, voglibose, pradimicin-Q, salbostatin, CKD-711, MDL-25,637, MDL-73,945, and MOR 14; (7) alpha-amylase inhibitors such as tendamistat, trestatin, and AI-3688; (8) insulin secretagogues such as liinogliride and A4166; (9) fatty acid oxidation inhibitors such as clomoxir and etomoxir; (10) adenosine A2 antagonists such as midaglizole, isaglidole, deriglidole, idazoxan, earoxan, and fluparoxan; (11) insulin or insulin mimetics such as biota, LP-100, novarapid, insulin detemir, insulin lispro, insulin glargine, insulin zinc suspension (lente and ultralente), Lys-Pro insulin, GLP-1 (73-7) (insulintropin), and GLP-1 (7-36)-NH₂; (12) non-thiazolidinediones such as JT-501 and farglitazar (GW-2570/G1-262579); (13) PPAR α/γ dual agonists such as BVT-142, CLX-0940, GW-1536, GW1929, GW-2433, KRP-297, L-796449, LR-90, MK-0767, SB 219994, muragliptazar and reglitazar (ITT-501); (14) other insulin sensitizing drugs; (15) VPAC2 receptor agonists; (16) GLK modulators such as those disclosed in WO 03/015774; (17) retinoid modulators such as those disclosed in WO 03/000249; (18) GSK 3beta/GSK 3 inhibitors such as 4-[2-(2-bromophenyl)-4-(4-fluorophenyl)-1H-imidazol-5-yl]pyridine; (19) glycogen phosphorylase (HGLPa) inhibitors such as those disclosed in WO 03/037864; (20) ATP consumption promoters such as those disclosed in WO 03/007990; (21) TRB3 inhibitors, (22) vanilloid receptor ligands such as those disclosed in WO 03/049702, (23) hypoglycemic agents such as those disclosed in WO 03/015781 and WO 03/040114; and (24) Insulin-responsive DNA binding protein-1 (IRDBP-1) as disclosed in WO 03/057827.

V. Stimulation of Glucose Production

[0186] Decreases in SOGA polypeptide levels and/or activity result in the stimulation of glucose production in cells. Thus, inhibitors of the SOGA polypeptides and polynucle-

otides of the invention can be used in methods in which an increase in glucose production is desired for research, diagnostic, and/or therapeutic proposes. These methods can be carried using techniques to decrease the expression and/or activity of SOGA polypeptides in a cell, in a tissue, and/or in a subject,

[0187] One aspect of the invention relates to a method of increasing glucose production in a cell, comprising contacting said cell with an agent that decreases the activity of a polynucleotide or polypeptide of the invention in an amount effective to increase glucose production in the cell.

[0188] Another aspect of the invention relates to a method of increasing autophagy in a cell, comprising contacting said cell with an agent that decreases the activity of a polynucleotide or polypeptide of the invention in an amount effective to increase autophagy in said cell.

[0189] The cells to be contacted can be in vitro, ex vivo, or in vivo (e.g., in an animal model of disease or a patient). Cells can be contacted with an agent by any means known in the art and as described herein.

[0190] A further aspect of the invention relates to a method of increasing blood glucose levels in a subject, comprising delivering to said subject an agent that decreases the activity of a polynucleotide or polypeptide of the invention in an amount effective to increase the blood glucose levels in said subject.

[0191] Another aspect of the invention relates to a method of decreasing insulin sensitivity in a subject, comprising delivering to said subject an agent that decreases the activity of a polynucleotide or polypeptide of the invention in an amount effective to decrease insulin sensitivity in said subject.

[0192] In one embodiment, the subject is one that is in need of increased glucose levels and/or decreased insulin sensitivity. The subject can currently have or be at risk for a carbohydrate-related metabolic disorder such as hypoglycemia, e.g., as a result of sepsis, malaria, or injection of insulin.

[0193] Agents that can be used in the methods of the invention include, without limitation, an antisense oligonucleotide, ribozyme, or siRNA that targets a SOGA polynucleotide, an antibody or antibody fragment that binds to a SOGA polypeptide, agents that modulate the insulin and/or adiponectin signaling pathways, and agents identified by the screening methods described below.

[0194] The agents of the present invention can optionally be delivered in conjunction with other therapeutic agents. The additional therapeutic agents can be delivered concurrently with the agents of the invention.

VI. Monitoring of Responsiveness to Treatment

[0195] The increased levels of SOGA polypeptide in response to administration of insulin and adiponectin provides the basis for monitoring responsiveness of a subject to anti-diabetic treatments. It is known that insulin treatment of diabetics is not effective 100% of the time and that certain drugs may induce adiponectin but do not necessarily lower glucose. Measuring the induction of SOGA in response to an anti-diabetic treatment may provide insight into the ability of a subject to respond to the treatment and can help identify subjects that are likely to respond or not respond to a particular treatment.

[0196] One aspect of the invention relates to a method of measuring the response of a subject to a treatment for diabetes, comprising determining the circulating level of a SOGA

polypeptide or a functional fragment thereof in said subject after administration of the treatment and comparing it to the circulating level of the polypeptide or a functional fragment thereof in said subject before administration of the treatment.

[0197] Another aspect of the invention relates to a method of predicting the clinical outcome of a diabetes treatment in a subject, comprising determining the circulating level of a SOGA polypeptide or a functional fragment thereof in said subject after administration of the treatment and comparing it to the circulating level of the polypeptide or a functional fragment thereof in said subject before administration of the treatment.

[0198] In these methods, an increase in circulating levels of SOGA polypeptide or a functional fragment thereof subsequent to administration of an anti-diabetic treatment is indicative that the subject will respond to the treatment (e.g., the treatment will lower glucose levels). Conversely, if the circulating level of SOGA does not increase or increases less than a "normal" amount, the subject may not respond favorably to the treatment. The magnitude of the increase in SOGA polypeptide (e.g., a "normal" increase as compared to a "less than normal" increase in SOGA) can be classified based on average numbers in a population of similar subjects.

[0199] In one embodiment, determining the level of a SOGA polypeptide comprises determining the level the polypeptide. Determining the level of a polypeptide can be carried out by any means known in the art and as described herein, such as Western blots, immunoblots, immunoprecipitation, immunohistochemistry, immunofluorescence, enzyme-linked immunosorbant assays, and radioimmunoassays. Assays for expression and/or activity can be carried out automatically or partially automatically in a machine or apparatus designed to perform such assays, e.g., using computer-assisted methods. The results of the assays can be stored in a computer database and analyzed to produce predictive results. In some embodiments, the data can be analyzed, e.g., by comparing intra-patient results over time or before and after treatment or comparing inter-patient results to determine baseline and/or abnormal values in a population.

[0200] In a further embodiment, determining the level of a SOGA polypeptide comprises determining the activity of the polypeptides. The activity may be any activity associated with the polypeptide, including, without limitation, inhibition of glucose production, enzyme activity, protein interaction, receptor binding, ligand binding, a cell signal transduction event, etc.

[0201] In one embodiment, determining the level of a SOGA polypeptide comprises determining the level of a nucleic acid encoding the polypeptide. Determining the level of a nucleic acid can be carried out by any means known in the art and as described herein, such as Northern blots, dot blots, PCR, RT-PCR, quantitative PCR, sequence analysis, gene microarray analysis, in situ hybridization, and detection of a reporter gene.

[0202] One aspect of the invention relates to kits useful for carrying out the methods of the invention. One embodiment relates to kits for determining the level of expression and/or activity of SOGA, e.g., to assess responsiveness to anti-diabetic treatment, comprising a reagent for determining the expression and/or activity of a SOGA polypeptide or a functional fragment thereof. The reagents may be nucleic acids (e.g., an oligonucleotide that specifically hybridizes to a nucleic acid encoding a SOGA polypeptide and can be used as a hybridization probe or an amplification primer), antibod-

ies (e.g., one that specifically binds to a SOGA polypeptide), or other agents that specifically recognize the polynucleotides or polypeptides of the invention.

[0203] The reagents can be conjugated to a detectable tag or detectable label. Such a tag can be any suitable tag which allows for detection of the reagents and includes, but is not limited to, any composition or label detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads™), fluorescent dyes (e.g., fluorescein, Texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads.

[0204] In addition, the reagents can be immobilized on a substrate. Such a substrate can include any suitable substrate for immobilization of a detection reagent such as would be used in any of the previously described methods of detection. Briefly, a substrate suitable for immobilization of a detection reagent includes any solid support, such as any solid organic, biopolymer or inorganic support that can form a bond with the detection reagent without significantly effecting the activity and/or ability of the detection reagent to detect the desired target molecule. Exemplary organic solid supports include polymers such as polystyrene, nylon, phenol-formaldehyde resins, acrylic copolymers (e.g., polyacrylamide), stabilized intact whole cells, and stabilized crude whole cell/membrane homogenates. Exemplary biopolymer supports include cellulose, polydextrans (e.g., Sephadex®), agarose, collagen and chitin. Exemplary inorganic supports include glass beads (porous and nonporous), stainless steel, metal oxides (e.g., porous ceramics such as ZrO₂, TiO₂, Al₂O₃, and NiO) and sand.

[0205] The kits may further comprise other components useful for detecting expression or activity, e.g., buffers, cells, culture medium, enzymes, labeling reagents, containers, etc.

[0206] In one embodiment, the kit comprises an array of reagents for determining expression and/or activity. The array can comprise a substrate having a plurality of addresses. At least one address of the plurality includes a capture probe that binds specifically to a polynucleotide or polypeptide of the invention. The array can have a density of at least, or less than, 10, 20 50, 100, 200, 500, 700, 1,000, 2,000, 5,000 or 10,000 or more addresses/cm², and ranges between. The substrate can be a two-dimensional substrate such as a glass slide, a wafer (e.g., silica or plastic), a mass spectroscopy plate, or a three-dimensional substrate such as a gel pad. Addresses in addition to addresses of the plurality can be disposed on the array.

[0207] In one embodiment, at least one address of the plurality includes a nucleic acid capture probe that hybridizes specifically to a polynucleotide of the invention, e.g., the sense or anti-sense strand. Each address of the subset can include a capture probe that hybridizes to a different region of a polynucleotide. An array can be generated by any of a variety of methods. Appropriate methods include, e.g., photolithographic methods (e.g., U.S. Pat. Nos. 5,143,854; 5,510,270; and 5,527,681), mechanical methods (e.g., directed-flow methods as described in U.S. Pat. No. 5,384,

261), pin-based methods (e.g., as described in U.S. Pat. No. 5,288,514), and bead-based techniques (e.g., as described in PCT US/93/04145).

[0208] In another embodiment, at least one address of the plurality includes a polypeptide capture probe that binds specifically to a polypeptide of the invention or fragment thereof. The polypeptide capture probe can be a naturally-occurring interaction partner of a SOGA polypeptide. In one embodiment, the polypeptide is an antibody, e.g., an antibody specific for a SOGA polypeptide, such as a polyclonal antibody, a monoclonal antibody, or a single-chain antibody.

VII. Screening Assays and Animal Models

[0209] The identification of polynucleotides and polypeptides that are involved in insulin and adiponectin signaling and glucose regulation provides targets that can be used to screen for agents that regulate glucose production as well as models for studying these pathways in vitro or in animals.

[0210] One aspect of the invention relates to a method of identifying an agent that binds to a SOGA polypeptide or a functional fragment thereof of the invention, comprising:

[0211] contacting the polypeptide or a functional fragment thereof with a test agent under conditions whereby binding between the polypeptide or a functional fragment thereof and the test agent can occur; and

[0212] detecting binding between the polypeptide or a functional fragment thereof and the test agent.

[0213] Another aspect of the invention relates to a method of identifying an agent that modulates the activity of a SOGA polypeptide or a functional fragment thereof of the invention, comprising:

[0214] contacting the polypeptide or a functional fragment thereof with a test agent under conditions whereby modulation of the activity of the polypeptide or a functional fragment thereof can occur; and

[0215] detecting modulation of the activity of the polypeptide or a functional fragment thereof upon contact with the test agent as compared to activity of the polypeptide or a functional fragment thereof in the absence of contact with the test agent.

[0216] In each aspect above, the assay may be a cell-based or cell-free assay. In one embodiment, the cell may be a primary cell, e.g., an endothelial cell or a tumor cell, such as a breast tumor cell. In another embodiment, the cell is from a cell line, e.g., a hepatocyte, kidney, or muscle cell line or a tumor cell line. The cell may be contacted with the agent in vitro (e.g., in a culture dish) or in an animal (e.g., a transgenic animal or an animal model). In one embodiment, the detected increase or decrease in expression and/or activity is statistically significant, e.g., at least p<0.05, e.g., p<0.01, 0.005, or 0.001. In another embodiment, the detected increase or decrease is at least about 10%, 20%, 30%, 40%, 50%, 60&, 70%, 80%, 90%, 100% or more.

[0217] Any desired end-point can be detected in a screening assay, e.g., binding to the polypeptide, gene or RNA, modulation of the activity of the polypeptide, modulation of glucose-related pathways, and/or interference with binding by a known regulator of a polynucleotide or polypeptide. Methods of detecting the foregoing activities are known in the art and include the methods disclosed herein.

[0218] Any agent of interest can be screened according to the present invention. Suitable test agents include organic and inorganic molecules. Suitable organic molecules can include but are not limited to small molecules (compounds less than

about 1000 Daltons), polypeptides (including enzymes, antibodies, and Fab' fragments), carbohydrates, lipids, coenzymes, and nucleic acid molecules (including DNA, RNA, and chimerics and analogs thereof) and nucleotides and nucleotide analogs. In particular embodiments, the agent is an antisense nucleic acid, an siRNA, or a ribozyme that inhibits production of a SOGA polypeptide.

[0219] Further, the methods of the invention can be practiced to screen an agent library, e.g., a small molecule library, a combinatorial chemical compound library, a polypeptide library, a cDNA library, a library of antisense nucleic acids, and the like, or an arrayed collection of agents such as polypeptide and nucleic acid arrays.

[0220] In one representative embodiment, the invention provides methods of screening test agents to identify a test agent that binds to a SOGA polypeptide or functional fragment thereof. Agents that are identified as binding to the polypeptide or functional fragment can be subject to further screening (e.g., for modulation of glucose production) using the methods described herein or other suitable techniques.

[0221] Also provided are methods of screening agents to identify those that modulate the activity of a SOGA polypeptide or functional fragment thereof. The term "modulate" is intended to refer to agents that enhance (e.g., increase) or inhibit (e.g., reduce) the activity of the polypeptide (or functional fragment). For example, the interaction of the polypeptide or functional fragment with a binding partner can be evaluated. As another alternative, physical methods, such as NMR, can be used to assess biological function. Activity of the SOGA polypeptides or functional fragment can be evaluated by any method known in the art, including the methods disclosed herein.

[0222] Agents that are identified as modulators of activity can optionally be further screened using the methods described herein (e.g., for binding to the SOGA polypeptide or functional fragment thereof, polynucleotide or RNA, modulation of glucose, and the like). The agent can directly interact with the polypeptide or functional fragment, polynucleotide or mRNA and thereby modulate its activity. Alternatively, the agent can interact with any other polypeptide, nucleic acid or other molecule as long as the interaction results in a modulation of the activity of the SOGA polypeptide or functional fragment.

[0223] With respect to cell-free binding assays, test agents can be synthesized or otherwise affixed to a solid substrate, such as plastic pins, glass slides, plastic wells, and the like. For example, the test agents can be immobilized utilizing conjugation of biotin and streptavidin by techniques well known in the art. The test agents are contacted with the polypeptide or functional fragment thereof and washed. Bound polypeptide can be detected using standard techniques in the art (e.g., by radioactive or fluorescence labeling of the polypeptide or functional fragment, by ELISA methods, and the like).

[0224] Alternatively, the target can be immobilized to a solid substrate and the test agents contacted with the bound polypeptide or functional fragment thereof. Identifying those test agents that bind to and/or modulate the SOGA polypeptide or functional fragment can be carried out with routine techniques. For example, the test agents can be immobilized utilizing conjugation of biotin and streptavidin by techniques well known in the art. As another illustrative example, antibodies reactive with the polypeptide or functional fragment can be bound to the wells of the plate, and the polypeptide

trapped in the wells by antibody conjugation. Preparations of test agents can be incubated in the polypeptide (or functional fragment)-presenting wells and the amount of complex trapped in the well can be quantitated.

[0225] In another representative embodiment, a fusion protein can be provided which comprises a domain that facilitates binding of the polypeptide to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtiter plates, which are then combined with cell lysates (e.g., 35 S-labeled) and the test agent, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel detected directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of SOGA polypeptide or functional fragment thereof found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

[0226] Another technique for agent screening provides for high throughput screening of agents having suitable binding affinity to the polypeptide of interest, as described in published PCT application WO84/03564. In this method, a large number of different small test agents are synthesized on a solid substrate, such as plastic pins or some other surface. The test agents are reacted with the SOGA polypeptide or functional fragment thereof and washed. Bound polypeptide is then detected by methods well known in the art. Purified polypeptide or a functional fragment can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

[0227] With respect to cell-based assays, any suitable cell can be used, including bacteria, yeast, insect cells (e.g., with a baculovirus expression system), avian cells, mammalian cells, or plant cells. In exemplary embodiments, the assay is carried out in a cell line that naturally expresses the polynucleotide or produces the polypeptide, e.g., hepatocytes or renal cells. Further, in other embodiments, it is desirable to use nontransformed cells (e.g., primary cells) as transformation may alter the function of the polypeptide.

[0228] The screening assay can be used to detect agents that bind to or modulate the activity of the native SOGA polypeptide (e.g., polypeptide that is normally produced by the cell). Alternatively, the cell can be modified to express (e.g., over-express) a recombinant SOGA polypeptide or functional fragment thereof. According to this embodiment, the cell can be transiently or stably transformed with a polynucleotide encoding the SOGA polypeptide or functional fragment, but is preferably stably transformed, for example, by stable integration into the genome of the organism or by expression from a stably maintained episome (e.g., Epstein Barr Virus derived episomes). In another embodiment, a polynucleotide encoding a reporter molecule can be linked to a regulatory element of the polynucleotide encoding a SOGA polypeptide and used to identify compounds that modulate expression of the polypeptide.

[0229] In a cell-based assay, the agent to be screened can interact directly with the SOGA polypeptide or functional fragment thereof (i.e., bind to it) and modulate the activity thereof. Alternatively, the agent can be one that modulates

polypeptide activity (or the activity of a functional fragment) at the nucleic acid level. To illustrate, the agent can modulate transcription of the gene (or transgene), modulate the accumulation of mRNA (e.g., by affecting the rate of transcription and/or turnover of the mRNA), and/or modulate the rate and/or amount of translation of the mRNA transcript.

[0230] As a further type of cell-based binding assay, the SOGA polypeptide or functional fragment thereof can be used as a "bait protein" in a two-hybrid or three-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al., *Cell* 72:223 (1993); Madura et al., *J. Biol. Chem.* 268:12046 (1993); Bartel et al., *Biotechniques* 14:920 (1993); Iwabuchi et al., *Oncogene* 8:1693 (1993); and PCT publication WO94/10300), to identify other polypeptides that bind to or interact with the polypeptide of the invention or functional fragment thereof.

[0231] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the polynucleotide that encodes the SOGA polypeptide or functional fragment thereof is fused to a nucleic acid encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, optionally from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a nucleic acid that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact in vivo, forming a complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter sequence (e.g., LacZ), which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the nucleic acid encoding the polypeptide that exhibited binding to the SOGA polypeptide or functional fragment.

[0232] As another cell-based assay, the invention provides a method of screening an agent for modulation of glucose production. In particular embodiments, the cell comprises an isolated polynucleotide encoding the SOGA polypeptide or functional fragment thereof. According to this embodiment, it is preferred that the isolated polynucleotide encoding the polypeptide or functional fragment is stably incorporated into the cell (i.e., by stable integration into the genome of the organism or by expression from a stably maintained episome such as Epstein Barr Virus derived episomes).

[0233] Screening assays can also be carried out in vivo in animals. Thus, as still a further aspect, the invention provides a transgenic non-human animal comprising an isolated polynucleotide encoding a SOGA polypeptide or functional fragment thereof, which can be produced according to methods well-known in the art. The transgenic non-human animal can be from any species, including avians and non-human mammals. According to this aspect of the invention, suitable non-human mammals include mice, rats, rabbits, guinea pigs, goats, sheep, pigs, and cattle. Suitable avians include chickens, ducks, geese, quail, turkeys, and pheasants.

[0234] The polynucleotide encoding the polypeptide or functional fragment can be stably incorporated into cells within the transgenic animal (typically, by stable integration into the genome or by stably maintained episomal constructs). It is not necessary that every cell contain the transgene, and the animal can be a chimera of modified and

unmodified cells, as long as a sufficient number of cells comprise and express the polynucleotide encoding the polypeptide or functional fragment so that the animal is a useful screening tool.

[0235] Exemplary methods of using the transgenic non-human animals of the invention for in vivo screening of agents that modulate glucose production and/or the activity of a SOGA polypeptide comprise administering a test agent to a transgenic non-human animal (e.g., a mammal such as a mouse) comprising an isolated polynucleotide encoding a SOGA polypeptide or functional fragment thereof stably incorporated into the genome and detecting whether the test agent modulates glucose levels and/or polypeptide activity (or the activity of a functional fragment). It is known in the art how to measure these responses in vivo.

[0236] Methods of making transgenic animals are known in the art. DNA or RNA constructs can be introduced into the germ line of an avian or mammal to make a transgenic animal. For example, one or several copies of the construct can be incorporated into the genome of an embryo by standard transgenic techniques.

[0237] In an exemplary embodiment, a transgenic non-human animal is produced by introducing a transgene into the germ line of the non-human animal. Transgenes can be introduced into embryonal target cells at various developmental stages. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used should, if possible, be selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness.

[0238] Introduction of the transgene into the embryo can be accomplished by any of a variety of means known in the art such as microinjection, electroporation, lipofection, or a viral vector. For example, the transgene can be introduced into a mammal by microinjection of the construct into the pronuclei of the fertilized mammalian egg(s) to cause one or more copies of the construct to be retained in the cells of the developing mammal(s). Following introduction of the transgene construct into the fertilized egg, the egg can be incubated in vitro for varying amounts of time, or reimplanted into the surrogate host, or both. One common method is to incubate the embryos in vitro for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

[0239] The progeny of the transgenically manipulated embryos can be tested for the presence of the construct by Southern blot analysis of a segment of tissue. An embryo having one or more copies of the exogenous cloned construct stably integrated into the genome can be used to establish a permanent transgenic animal line.

[0240] Transgenically altered animals can be assayed after birth for the incorporation of the construct into the genome of the offspring. This can be done by hybridizing a probe corresponding to the polynucleotide sequence coding for the polypeptide or a segment thereof onto chromosomal material from the progeny. Those progeny found to contain at least one copy of the construct in their genome are grown to maturity.

[0241] Methods of producing transgenic avians are also known in the art, see, e.g., U.S. Pat. No. 5,162,215.

[0242] In particular embodiments, to create an animal model in which the activity or expression of a SOGA polypeptide is decreased, it is desirable to inactivate, replace or knock-out the endogenous gene encoding the polypeptide by homologous recombination with a transgene using embryonic stem cells. In this context, a transgene is meant to refer to

heterologous nucleic acid that upon insertion within or adjacent to the gene results in a decrease or inactivation of gene expression or polypeptide amount or activity.

[0243] A knock-out of a gene means an alteration in the sequence of a gene that results in a decrease of function of the gene, preferably such that the gene expression or polypeptide amount or activity is undetectable or insignificant. Knock-outs as used herein also include conditional knock-outs, where alteration of the gene can occur upon, for example, exposure of the animal to a substance that promotes gene alteration (e.g., tetracycline or ecdysone), introduction of an enzyme that promotes recombination at a gene site (e.g., Cre in the Cre-lox system), or other method for directing the gene alteration postnatally. Knock-out animals may be prepared using methods known to those of skill in the art. See, for example, Hogan, et al. (1986) Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

[0244] A knock-out construct is a nucleic acid sequence, such as a DNA or RNA construct, which, when introduced into a cell, results in suppression (partial or complete) of expression of a polypeptide encoded by endogenous DNA in the cell. A knock-out construct as used herein may include a construct containing a first fragment from the 5' end of the gene encoding a SOGA polypeptide, a second fragment from the 3' end of the gene and a DNA fragment encoding a selectable marker positioned between the first and second fragments. It should be understood by the skilled artisan that any suitable 5' and 3' fragments of a gene may be used as long as the expression of the corresponding gene is partially or completely suppressed by insertion of the transgene. Suitable selectable markers include, but are not limited to, neomycin, puromycin and hygromycin. In addition, the construct may contain a marker, such as diphtheria toxin A or thymidine kinase, for increasing the frequency of obtaining correctly targeted cells. Suitable vectors include, but are not limited to, pBLUESCRIPT, pBR322, and pGEM7.

[0245] Alternatively, a knock-out construct may contain RNA molecules such as antisense RNA, siRNA, and the like to decrease the expression of a gene encoding a SOGA polypeptide. Typically, for stable expression the RNA molecule is placed under the control of a promoter. The promoter may be regulated, if deficiencies in the protein of interest may lead to a lethal phenotype, or the promoter may drive constitutive expression of the RNA molecule such that the gene of interest is silenced under all conditions of growth. While homologous recombination between the knock-out construct and the gene of interest may not be necessary when using an RNA molecule to decrease gene expression, it may be advantageous to target the knock-out construct to a particular location in the genome of the host organism so that unintended phenotypes are not generated by random insertion of the knock-out construct.

[0246] The knock-out construct may subsequently be incorporated into a viral or nonviral vector for delivery to the host animal or may be introduced into embryonic stem (ES) cells. ES cells are typically selected for their ability to integrate into and become part of the germ line of a developing embryo so as to create germ line transmission of the knock-out construct. Thus, any ES cell line that can do so is suitable for use herein. Suitable cell lines which may be used include, but are not limited to, the 129J ES cell line or the J1 ES cell line. The cells are cultured and prepared for DNA insertion using methods well-known to the skilled artisan (e.g., see

Robertson (1987) In: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E. J. Robertson, ed. IRL Press, Washington, D.C.; Bradley et al., *Curr. Topics Develop. Biol.* 20:357 (1986); Hogan et al., (1986) Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

[0247] Insertion of the knock-out construct into the ES cells may be accomplished using a variety of methods well-known in the art, including, for example, electroporation, microinjection, and calcium phosphate treatment. For insertion of the DNA or RNA sequence, the knock-out construct nucleic acids are added to the ES cells under appropriate conditions for the insertion method chosen. If the cells are to be electroporated, the ES cells and construct nucleic acids are exposed to an electric pulse using an electroporation machine (electroporator) and following the manufacturer's guidelines for use. After electroporation, the cells are allowed to recover under suitable incubation conditions. The cells are then screened for the presence of the knockout construct.

[0248] Each knock-out construct to be introduced into the cell is first typically linearized if the knock-out construct has been inserted into a vector. Linearization is accomplished by digesting the knock-out construct with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knock-out construct sequence.

[0249] Screening for cells which contain the knock-out construct (homologous recombinants) may be done using a variety of methods. For example, as described herein, cells can be processed as needed to render DNA in them available for hybridization with a nucleic acid probe designed to hybridize only to cells containing the construct. For example, cellular DNA can be probed with ³²P-labeled DNA which locates outside the targeting fragment. This technique can be used to identify those cells with proper integration of the knock-out construct. The DNA can be extracted from the cells using standard methods (e.g., see, Sambrook et al., *Molecular Cloning: A Laboratory Manual* 2nd Ed. (Cold Spring Harbor, N.Y., 1989)). The DNA may then be analyzed by Southern blot with a probe or probes designed to hybridize in a specific pattern to genomic DNA digested with one or more particular restriction enzymes.

[0250] Once appropriate ES cells are identified, they are introduced into an embryo using standard methods. They can be introduced using microinjection, for example. Embryos at the proper stage of development for integration of the ES cell to occur are obtained, such as by perfusion of the uterus of pregnant females. For example, mouse embryos at 3-4 days development can be obtained and injected with ES cells using a micropipet. After introduction of the ES cell into the embryo, the embryo is introduced into the uterus of a pseudopregnant female mouse. The stage of the pseudopregnancy is selected to enhance the chance of successful implantation. In mice, 2-3 days pseudopregnant females are appropriate.

[0251] Germline transmission of the knockout construct may be determined using standard methods. Offspring resulting from implantation of embryos containing the ES cells described above are screened for the presence of the desired alteration (e.g., knock-out of the SOGA polypeptide). This may be done, for example, by obtaining DNA from offspring (e.g., tail DNA) to assess for the knock-out construct, using known methods (e.g., Southern analysis, dot blot analysis, PCR analysis). See, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual* 2nd Ed. (Cold Spring Harbor,

N.Y., 1989). Offspring identified as chimeras may be crossed with one another to produce homozygous knock-out animals.

[0252] Mice are often used as animal models because they are easy to house, relatively inexpensive, and easy to breed. However, other knock-out animals may also be made in accordance with the present invention such as, but not limited to, monkeys, cattle, sheep, pigs, goats, horses, dogs, cats, guinea pigs, rabbits and rats. Accordingly, appropriate vectors and promoters well-known in the art may be selected and used to generate a transgenic animal deficient in expression of a SOGA polypeptide.

[0253] In another embodiment, animal models may be created using animals that are not transgenic. For example, animal models of diabetes or obesity are well known in the art and can be used to study the effects of regulators of glucose production.

VIII. Pharmaceutical Compositions

[0254] As a further aspect, the invention provides pharmaceutical formulations and methods of administering the same to achieve any of the diagnostic or therapeutic effects (e.g., inhibition or stimulation of glucose production) discussed above. The pharmaceutical formulation may comprise any of the reagents discussed above in a pharmaceutically acceptable carrier, e.g., a polynucleotide encoding a SOGA polypeptide or a fragment thereof or a vector or cell comprising the polynucleotide, a SOGA polypeptide or fragment thereof, an antibody against a SOGA polypeptide, an anti-sense oligonucleotide, an siRNA molecule, a ribozyme, an aptamer, a peptidomimetic, a small molecule, or any other agent that modulates the activity of a SOGA polypeptide, including agents identified by the screening methods described herein.

[0255] By "pharmaceutically acceptable" it is meant a material that is not biologically or otherwise undesirable, i.e., the material can be administered to a subject without causing any undesirable biological effects such as toxicity.

[0256] The formulations of the invention can optionally comprise medicinal agents, pharmaceutical agents, carriers, adjuvants, dispersing agents, diluents, and the like.

[0257] The agents of the invention can be formulated for administration in a pharmaceutical carrier in accordance with known techniques. See, e.g., Remington, *The Science And Practice of Pharmacy* (9th Ed. 1995). In the manufacture of a pharmaceutical formulation according to the invention, the agent (including the physiologically acceptable salts thereof) is typically admixed with, inter alia, an acceptable carrier. The carrier can be a solid or a liquid, or both, and is preferably formulated with the agent as a unit-dose formulation, for example, a tablet, which can contain from 0.01 or 0.5% to 95% or 99% by weight of the agent. One or more agents can be incorporated in the formulations of the invention, which can be prepared by any of the well-known techniques of pharmacy.

[0258] A further aspect of the invention is a method of treating subjects in vivo, comprising administering to a subject a pharmaceutical composition comprising an agent of the invention in a pharmaceutically acceptable carrier, wherein the pharmaceutical composition is administered in a therapeutically effective amount. Administration of the compounds of the present invention to a human subject or an animal in need thereof can be by any means known in the art for administering agents.

[0259] The formulations of the invention include those suitable for oral, rectal, topical, buccal (e.g., sub-lingual), vaginal, parenteral (e.g., subcutaneous, intramuscular including skeletal muscle, cardiac muscle, diaphragm muscle and smooth muscle, intradermal, intravenous, intraperitoneal), topical (i.e., both skin and mucosal surfaces, including airway surfaces), intranasal, transdermal, intraarticular, intrathecal, and inhalation administration, administration to the liver by intraportal delivery, as well as direct organ injection (e.g., into the liver, kidney or muscle). The most suitable route in any given case will depend on the nature and severity of the condition being treated and on the nature of the particular agent which is being used.

[0260] For injection, the carrier will typically be a liquid, such as sterile pyrogen-free water, pyrogen-free phosphate-buffered saline solution, bacteriostatic water, or Cremophor EL[R] (BASF, Parsippany, N.J.). For other methods of administration, the carrier can be either solid or liquid.

[0261] For oral administration, the agent can be administered in solid dosage forms, such as capsules, tablets, and powders, or in liquid dosage forms, such as elixirs, syrups, and suspensions. Agents can be encapsulated in gelatin capsules together with inactive ingredients and powdered carriers, such as glucose, lactose, sucrose, mannitol, starch, cellulose or cellulose derivatives, magnesium stearate, stearic acid, sodium saccharin, talcum, magnesium carbonate and the like. Examples of additional inactive ingredients that can be added to provide desirable color, taste, stability, buffering capacity, dispersion or other known desirable features are red iron oxide, silica gel, sodium lauryl sulfate, titanium dioxide, edible white ink and the like. Similar diluents can be used to make compressed tablets. Both tablets and capsules can be manufactured as sustained release products to provide for continuous release of medication over a period of hours. Compressed tablets can be sugar coated or film coated to mask any unpleasant taste and protect the tablet from the atmosphere, or enteric-coated for selective disintegration in the gastrointestinal tract. Liquid dosage forms for oral administration can contain coloring and flavoring to increase patient acceptance.

[0262] Formulations suitable for buccal (sub-lingual) administration include lozenges comprising the agent in a flavored base, usually sucrose and acacia or tragacanth; and pastilles comprising the agent in an inert base such as gelatin and glycerin or sucrose and acacia.

[0263] Formulations of the present invention suitable for parenteral administration comprise sterile aqueous and non-aqueous injection solutions of the agent, which preparations are preferably isotonic with the blood of the intended recipient. These preparations can contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the intended recipient. Aqueous and non-aqueous sterile suspensions can include suspending agents and thickening agents. The formulations can be presented in unit/dose or multi-dose containers, for example sealed ampoules and vials, and can be stored in a freeze-dried (lyophilized) condition requiring only the addition of the sterile liquid carrier, for example, saline or water-for-injection immediately prior to use.

[0264] Extemporaneous injection solutions and suspensions can be prepared from sterile powders, granules and tablets of the kind previously described. For example, in one aspect of the present invention, there is provided an injectable, stable, sterile composition comprising an agent of the

invention, in a unit dosage form in a sealed container. The agent or salt is provided in the form of a lyophilizate which is capable of being reconstituted with a suitable pharmaceutically acceptable carrier to form a liquid composition suitable for injection thereof into a subject. The unit dosage form typically comprises from about 1 mg to about 10 grams of the agent or salt. When the agent or salt is substantially water-insoluble, a sufficient amount of emulsifying agent which is pharmaceutically acceptable can be employed in sufficient quantity to emulsify the agent or salt in an aqueous carrier. One such useful emulsifying agent is phosphatidyl choline.

[0265] Formulations suitable for rectal administration are preferably presented as unit dose suppositories. These can be prepared by admixing the agent with one or more conventional solid carriers, for example, cocoa butter, and then shaping the resulting mixture.

[0266] Formulations suitable for topical application to the skin preferably take the form of an ointment, cream, lotion, paste, gel, spray, aerosol, or oil. Carriers which can be used include petroleum jelly, lanoline, polyethylene glycols, alcohols, transdermal enhancers, and combinations of two or more thereof.

[0267] Formulations suitable for transdermal administration can be presented as discrete patches adapted to remain in intimate contact with the epidermis of the recipient for a prolonged period of time. Formulations suitable for transdermal administration can also be delivered by iontophoresis (see, for example, Tyle, *Pharm. Res.* 3:318 (1986)) and typically take the form of an optionally buffered aqueous solution of the agent. Suitable formulations comprise citrate or bis/tris buffer (pH 6) or ethanol/water and contain from 0.1 to 0.2M of the agent.

[0268] The agent can alternatively be formulated for nasal administration or otherwise administered to the lungs of a subject by any suitable means, e.g., administered by an aerosol suspension of respirable particles comprising the agent, which the subject inhales. The respirable particles can be liquid or solid. The term "aerosol" includes any gas-borne suspended phase, which is capable of being inhaled into the bronchioles or nasal passages. Specifically, aerosol includes a gas-borne suspension of droplets, as can be produced in a metered dose inhaler or nebulizer, or in a mist sprayer. Aerosol also includes a dry powder composition suspended in air or other carrier gas, which can be delivered by insufflation from an inhaler device, for example. See Ganderton & Jones, *Drug Delivery to the Respiratory Tract*, Ellis Horwood (1987); Gonda (1990) *Critical Reviews in Therapeutic Drug Carrier Systems* 6:273-313; and Raeburn et al., *J. Pharmacol. Toxicol. Meth.* 27:143 (1992). Aerosols of liquid particles comprising the agent can be produced by any suitable means, such as with a pressure-driven aerosol nebulizer or an ultrasonic nebulizer, as is known to those of skill in the art. See, e.g., U.S. Pat. No. 4,501,729. Aerosols of solid particles comprising the agent can likewise be produced with any solid particulate medicament aerosol generator, by techniques known in the pharmaceutical art.

[0269] Alternatively, one can administer the agent in a local rather than systemic manner, for example, in a depot or sustained-release formulation.

[0270] Further, the present invention provides liposomal formulations of the agents disclosed herein and salts thereof. The technology for forming liposomal suspensions is well known in the art. When the agent or salt thereof is an aqueous-soluble salt, using conventional liposome technology, the

same can be incorporated into lipid vesicles. In such an instance, due to the water solubility of the agent or salt, the agent or salt will be substantially entrained within the hydrophilic center or core of the liposomes. The lipid layer employed can be of any conventional composition and can either contain cholesterol or can be cholesterol-free. When the agent or salt of interest is water-insoluble, again employing conventional liposome formation technology, the salt can be substantially entrained within the hydrophobic lipid bilayer which forms the structure of the liposome. In either instance, the liposomes which are produced can be reduced in size, as through the use of standard sonication and homogenization techniques.

[0271] The liposomal formulations containing the agents disclosed herein or salts thereof, can be lyophilized to produce a lyophilizate which can be reconstituted with a pharmaceutically acceptable carrier, such as water, to regenerate a liposomal suspension.

[0272] In the case of water-insoluble agent s, a pharmaceutical composition can be prepared containing the water-insoluble agent, such as for example, in an aqueous base emulsion. In such an instance, the composition will contain a sufficient amount of pharmaceutically acceptable emulsifying agent to emulsify the desired amount of the agent. Particularly useful emulsifying agents include phosphatidyl choline and lecithin,

[0273] In particular embodiments, the agent is administered to the subject in a therapeutically effective amount, as that term is defined above. Dosages of pharmaceutically active agents can be determined by methods known in the art, see, e.g., *Remington's Pharmaceutical Sciences* (Maack Publishing Co., Easton, Pa.). The therapeutically effective dosage of any specific agent will vary somewhat from agent to agent, and patient to patient, and will depend upon the condition of the patient and the route of delivery. As a general proposition, a dosage from about 0.1 to about 50 mg/kg will have therapeutic efficacy, with all weights being calculated based upon the weight of the agent, including the cases where a salt is employed. Toxicity concerns at the higher level can restrict intravenous dosages to a lower level such as up to about 10 mg/kg, with all weights being calculated based upon the weight of the agent, including the cases where a salt is employed. A dosage from about 10 mg/kg to about 50 mg/kg can be employed for oral administration. Typically, a dosage from about 0.5 mg/kg to 5 mg/kg can be employed for intramuscular injection. Particular dosages are about 1 mg/kg to 50 μmol/kg, and more particularly to about 22 μmol/kg and to 33 μmol/kg of the agent for intravenous or oral administration, respectively.

[0274] In particular embodiments of the invention, more than one administration (e.g., two, three, four, or more administrations) can be employed over a variety of time intervals (e.g., hourly, daily, weekly, monthly, etc.) to achieve therapeutic effects.

[0275] The present invention finds use in veterinary and medical applications. Suitable subjects include both avians and mammals, with mammals being preferred. The term "avian" as used herein includes, but is not limited to, chickens, ducks, geese, quail, turkeys, and pheasants. The term "mammal" as used herein includes, but is not limited to, humans, bovines, ovines, caprines, equines, felines, canines, lagomorphs, etc. Human subjects include neonates, infants, juveniles, and adults. In other embodiments, the subject is an animal model of diabetes or other metabolic disorder.

[0276] The present invention is more particularly described in the following examples that are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

Example 1

Identification of SOGA

[0277] Type II diabetes is associated with high glucose production. Obesity increases glucose production by lowering circulating levels of the hormone adiponectin. Therefore, type II diabetes can be treated by stimulating the adiponectin signaling pathway. Adiponectin lowers circulating glucose by inhibiting glucose production from the liver. Adiponectin inhibits glucose production by activating AMP-activated kinase (AMPK). AMPK stimulates fatty acid (FA) oxidation. The inhibition of glucose production by a signaling intermediate that increases FA oxidation is counter-intuitive because ATP generated from FA oxidation fuels glucose production. Furthermore, AMPK stimulates autophagy, a regulated mechanism of intracellular degradation that provides the biochemical intermediates for glucose production through the hydrolysis of proteins, glycogen and triglycerides. This deadlock led to the hypothesis that adiponectin inhibits glucose production through a novel mediator. Insulin inhibition of glucose production in the liver is mediated by the suppression of lysosome activity. We treated rat hepatoma cells with full-length recombinant adiponectin and identified the proteins that were bound to APPL1 in a co-immunoprecipitation assay using proteomics analysis. APPL1 was previously identified in a yeast 2-hybrid screen using the intracellular region of the adiponectin receptor. Proteomics analysis revealed a gene we are calling SOGA (also called TOA (Target Of Adiponectin)) that encodes a 161 kDa protein containing (1) a leucine zipper motif that enables binding to the leucine zipper motif of APPL1 and (2) Atg16 and Rab5-binding motifs that enable participation in membrane assembly for autophagy. The hydrolysis of proteins and glycogen by autophagy increases glucose production by producing biochemical intermediates for gluconeogenesis and glycogenolysis. Northern blot analysis revealed that SOGA is ubiquitously expressed as a 3.0 and a 4.5 kb mRNA. Our current hypothesis is that adiponectin stimulation of SOGA (NCBI Accession: FJ977045) can suppress glucose production.

[0278] We verified the expression of SOGA in the liver and other tissues by RT-PCR and Northern blot analysis. There are no publications describing SOGA, its gene, mRNA or amino acid sequence. The open reading frame of murine SOGA is derived from 16 exons. SOGA cDNA encodes a 1434 amino acid protein that lacks transmembrane domains. SOGA contains a leucine zipper motif that we predict allows SOGA to bind to the leucine zipper motif of APPL1 in our co-immunoprecipitation experiment (FIG. 1). The predicted regions of interest in SOGA include (1) a leucine zipper motif, (2) ATG16 motifs, (3) a Rab5 motif, (4) a casein kinase domain, (5) multiple myristylation and glycosylation sites and (6) multiple kinase specific phosphorylation sites (FIG. 1). Amino acid sequence alignment shows that murine SOGA is 91% identical to human SOGA. When substitutions for similar amino acids are taken into account, murine SOGA is 95% identical to human SOGA. SOGA is a highly conserved gene in mammals but absent in lower eukaryotes like yeast. Our current model is that adiponectin signaling triggers SOGA binding to APPL1, a proximal target of the adiponec-

tin receptor. Based on conserved domain predictions, SOGA binding to APPL1 contributes to adiponectin inhibition of protein degradation and glucose production. This may be accomplished through the binding of SOGA to APPL1, the proteolytic cleavage of SOGA and the secretion of its 25 kDa fragment.

[0279] The formation of the phagophore, a primary step in autophagy, can lead to the digestion of proteins and glycogen providing the biochemical intermediates for glucose production (FIG. 2). Atg16-Atg5-Atg12 forms a protein complex that is essential for the formation of an autophagosome. Atg12 is covalently conjugated to Atg5 by ubiquitination-like reactions that involve Atg7 and Atg10. Overexpression of Atg5 and Atg12 in yeast causes an increase in autophagy that is absent in mammalian cells, suggesting the existence of a novel protein in higher eukaryotes. Although 31 autophagy-related (Atg) proteins have been identified in yeast, SOGA is highly conserved in mammals but bears little homology to any gene product in yeast. Thus, the study of SOGA can lead to the elucidation of the mechanisms governing autophagy in mammals.

[0280] We predicted that SOGA plays a role in adiponectin's inhibition of glucose production based on its binding to APPL1 under adiponectin exposure and the conserved functional domains of SOGA which include (1) a leucine zipper motif that enables SOGA to bind to APPL1, (2) an ATG16 (autophagy 16) motif that enables SOGA to initiate autophagy through the formation of the phagophore, (3) Rab5 motif (a small GTPase) that enables the fusion of the autophagosome and lysosome, (4) casein kinase domain that enables a downstream signaling cascade, (5) myristylation and glycosylation sites that enable anchoring and (6) multiple kinase-specific phosphorylation sites that enable the modulation of SOGA by kinases and phosphatases (FIG. 1). Further insight into SOGA can increase our understanding of nutrient metabolism and lead to new ways of preventing and treating diabetes.

[0281] Species specific (mouse) SOGA peptide antigen (476) was detected with immune but not pre-immune sera from New Zealand White rabbits (FIG. 3, left panel). The signal intensity is proportional to the peptide antigen concentration. Using our rabbit polyclonal antisera (476) that is specific for mouse SOGA, SOGA was detected in mouse plasma at 25 kDa but not in human plasma (FIG. 3, right panel). Antisera from two different rabbits immunized with two different peptide antigens, 476 and 477 specific for mouse SOGA, detected a 25 kDa band in mouse plasma (FIG. 4). Antigen peptides 476 and 477 correspond to overlapping amino acid sequences in mouse SOGA.

[0282] The concentration of SOGA in plasma corresponded with circulating levels of adiponectin (FIG. 5). Plasma was sampled from young female C57Bl adiponectin null and wild-type mice. Western blot and densitometry of adiponectin and SOGA in ob/ob control mice and ob/ob mice treated with pioglitazone showed that adiponectin and SOGA were increased in ob/ob mice on pioglitazone compared to controls (FIG. 6). Western blot and densitometry of adiponectin and SOGA in ad libitum and calorie restricted fed C57 mice showed that adiponectin and SOGA were increased in calorie restricted mice compared to those fed ad libitum ($P < 0.05$ for statistical significance) (FIG. 7). Western blot and densitometry of adiponectin and SOGA in rapamycin and control fed C57Bl mice revealed that SOGA was decreased in rapamycin fed mice compared to controls ($P < 0.05$ for

statistical significance) (FIG. 8). FPLC fraction analysis of mouse plasma for SOGA was performed (FIG. 9). Graphs show SOGA, triglyceride, and cholesterol levels in FPLC fractions 11-33.

[0283] In summary, SOGA (TOA) is a novel protein that we have identified through proteomics and a co-immunoprecipitation assay; it binds to APPL1 under adiponectin exposure. The SOGA gene contains Atg16 and Rab5-binding motifs that are indicative of autophagic activities; it is hypothesized that adiponectin stimulation of SOGA can suppress glucose production. SOGA peptide antigen was detected by immune sera from NZW rabbits; SOGA was detected at 25 kDa in mouse plasma but not human plasma. Two distinct antigens corresponding to overlapping segments of SOGA produced antisera that detected a 25 kDa SOGA. Circulating levels of SOGA were greatly suppressed in adiponectin null (-/-) mice. Adiponectin and SOGA were increased by pioglitazone, and calorie restriction, but were suppressed by rapamycin. FPLC analysis indicates that SOGA circulates below 100 kDa.

Example 2

Experimental Methods

[0284] Mass Spectrometry. McArdle rat hepatoma cells were exposed to adipocyte conditioned media with or without adiponectin (Brooks et al., *J. Biol. Chem.* 282:35069 (2007)). Cell lysates were digested with proteomics grade trypsin (Sigma) and filtered through YM-10 molecular weight cutoff filters (Millipore, Bedford, Mass.). Tryptic digests were injected into an LCQ-Deca Ion Trap mass spectrometer coupled to a Surveyor HPLC system (Thermo Fisher Scientific, Waltham, Mass.). The solvent, 50% methanol and 0.1% formic acid, was delivered to the spectrometer at 200 μ L/min. Peptide masses were acquired in positive mode using electrospray ionization under the following source conditions: spray voltage was 5 kV, sheath gas was 40 (arbitrary units), auxiliary gas was 20 (arbitrary units), and heated capillary temperature was 350° C.

[0285] Cloning of Murine SOGA. Total RNA was obtained from primary mouse hepatocytes using Triazol reagent (Invitrogen). mRNA was isolated using Oligotex mRNA Kit (Qiagen). Primers used to clone SOGA were designed using publicly available genomic and mRNA sequence data based on the open reading frame of SOGA peptides detected by mass spectrometry. The 4.7 kb SOGA cDNA was isolated by annealing two PCR products using overlap extension. RNA ligase mediated RACE (Ambion) was used to clone the sequence from the 5'-end of SOGA mRNA. The cDNA for human SOGA was cloned by a similar method.

[0286] Antibody Production. Human- and murine-specific polyclonal antisera were produced in three New Zealand White rabbits (Franklin Rabbitry, NC) using a human-specific peptide antigen STQLTSFARSSRSAIRHSPSKC (SEQ ID NO:5) and two partially overlapping murine-specific peptide antigens CSAQSLASC*FIRPSRNPIRHSPSKC (SEQ ID NO:6) and SAQSLASC*FIRPSRNPIRHSPSKC (SEQ ID NO:7), where C* represents acemidomethyl cysteine. Synthetic peptides were purified by HPLC and analyzed on the LCQ-Deca Ion Trap mass spectrometer to confirm their molecular weight. Antigenic peptides (10 mg) were dissolved in 0.1 M NaH₂PO₄ (pH 7.2)/0.05 M NaCl and conjugated to keyhole limpet hemocyanin (KLH; 4 mg) before injection. KLH conjugated peptides were dissolved in 3 ml of 0.03%

trifluoroacetic acid and added to 3 ml complete Freund's adjuvant (Sigma). New Zealand White rabbits (Franklin Rabbitry, Wake Forest, N.C.) were injected intradermally using multiple injection sites. After 5 weeks, each animal was re-injected subcutaneously with KLH conjugated antigen in 1 ml of 50% incomplete Freund's adjuvant (Sigma). Four weeks later, 20 ml of blood were collected and rabbits were reimmunized. Injections and bleedings were performed at monthly intervals thereafter. The antibody production protocol was approved by UNC's Institutional Animal Care and Use Committee (IACUC).

[0287] Hepatocyte Studies. Mouse livers were perfused with a Krebs-Ringer-HEPES buffer containing collagenase (Sigma-Aldrich). Livers were isolated and cells were dispersed by gentle shaking and filtered through sterile nylon gauze. Cells were washed twice with sterile phosphate-buffered saline and purified by centrifugation in 50% isotonic Percoll (Sigma-Aldrich). Cells were resuspended with Krebs-Ringer-HEPES Ca²⁺ buffer to a total volume of 10 ml. Viability was validated via trypan blue exclusion and routinely exceeded 90%. Freshly isolated mouse hepatocytes were plated at 10⁵ cells per well in 12-well culture plates coated with rat tail collagen I (BD Biosciences). Cells were maintained in Dulbecco's modified Eagle medium (DMEM; Caisson Laboratories), 25 mM glucose and 10% horse serum (HS). Adiponectin was provided from adipocyte conditioned media with or without adiponectin (Brooks et al., *J. Biol. Chem.* 282:35069 (2007)). SOGA siRNA, AICAR (500 µM) or LY293004 (10 nM) were introduced to the media 48 hours before the measurement of glucose production. siRNA sequences corresponding to base pairs 333-351 and 1988-2007 on the open reading frame of murine SOGA were selected using a rational design algorithm (Invitrogen). Transfection with a pool of 2 siRNAs targeting SOGA had a greater knockdown efficiency than transfecting with the individual siRNAs. Transfection was achieved by electroporation using the Mouse Hepatocyte Nucleofector Kit (LONZA) according to the manufacturer's protocol. In brief, freshly isolated mouse hepatocytes were diluted to 3×10⁶ cells/tube in media without antibiotics and centrifuged at 2,000 rpm for 2 minutes. The supernatant was removed and the cells were resuspended in 100 µl of Nucleofector solution containing 100 nM of siRNA. The cell suspension was transferred to an electroporation cuvette which was placed in a Nucleofector I electroporation device and pulse charge was applied for 2 minutes using program T-28. Hepatocytes received 1.0 ml of media and were transferred to 12 well plates. SOGA expression, valine and glucose production were assayed 72 hours after siRNA transfection. Media was replaced with glucose-free DMEM containing MG-132 (10 µM), an inhibitor of the ubiquitin-proteasome pathway of protein degradation, for 6-8 hours to measure hepatocytes glucose production. Glucose was measured by colorimetric assay (Autokit Glucose CII) (Brooks et al., *J. Biol. Chem.* 282:35069 (2007)). Valine in the medium was measured by a HPLC (Waters) coupled TSQ-Quantum ultra triple quad mass analyzer (ThermoFinigan) in the Biomarkers Facility Core at UNC. Valine was measured in selected reaction monitoring mode (SRM) using the MS/MS transition of 118→72.

[0288] Lysosomal Activity. Autophagic activity was estimated by lysosome and late autophagosome vacuole staining using Lysotracker Red DND 99 (Invitrogen), a membrane permeable fluorescent labeled basic amine with high affinity for the acidic interior of the lysosome and late autophago-

some vacuole (Klionsky et al., *Autophagy* 4:151 (2008)). Cell medium was removed and replaced with GF/DMEM containing 50 nM Lysotracker Red. Cells were incubated for 30 min at 37° C. and the medium was replaced with GF/DMEM. Digital images were obtained at the Light Microscopy Facility at UNC with an Olympus 1×81 Motorized Inverted Microscope, a 40×11.30 Oil DIC lens, Camera pixel count: Hamamatsu C10600-1013 1344×10²⁴ using the acquisition software Velocity 5.3.2 (Perkin Elmer). Fluorescence Filter Cubes Specifications (Semrock, Inc.) were TXRED-4040B for rhodamine and Texas Red: Exciter 562 nm 20, Dichroic R 530-585/T 601-800, Emitter 642±20. Lysosome and late autophagosome vacuole number was determined from digital images as isolated punctate staining, greater than background staining intensity threshold, distinct from lipid droplets in clearly demarcated cells containing two nuclei. Spot recognition and enumeration according to the foregoing definition was determined by two individuals.

[0289] Mouse Studies. Mice were housed in ventilated isolator cage systems in a pathogen-free barrier facility maintained at 23° C., 55% humidity on a 12-h light/12-h dark cycle. Mice received a standard chow diet consisting of 73% carbohydrate, 18% protein, 4% fat and 5% ash (Purina). Young (3-6 month old) female C57B1/6J calorie restricted (CR) and ad libitum fed (AL) mice were maintained as previously described (Combs et al., *Diabetes* 52:268 (2003)). Adjustments were made to ensure that CR mice received 70% of the ad libitum food intake. Blood samples were collected at 1300 from the tail tip using heparinized capillary tubes (Fisher) and stored at -20° C. Male ob/ob mice (FVB background strain) received a daily dose of pioglitazone at 0.6 mg/kg BW in 0.025% (w/w) carboxymethylcellulose by oral gavage for 4 days. Control mice received carboxymethylcellulose by oral gavage for 4 days. Blood was collected from the tail tip on day 5 and analyzed for glucose, adiponectin and 25 kDa SOGA. Immediately after the collection of blood samples, ob/ob mice were sacrificed by cervical dislocation for tissue collection. Northern blot analysis for SOGA mRNA and 18S RNA was performed using 20 µg of liver RNA. NOD mice were bred and housed as previously described (Wong et al., *J. Immunol.* 176:1637 (2006)). Where indicated, diabetic NOD mice were injected with 5 units of insulin (NPH Human Insulin, Isophane Suspension; 100 U/ml Novolin; Novo Nordisk) 24 hours prior to blood collection. Adiponectin transgenic mice were produced as previously described (Combs et al., *Endocrinology* 145:367 (2004)). Glucose was measured by colorimetric assay. Adiponectin and SOGA were measured by SDS-PAGE analysis using 1 µl of plasma. The total concentration of protein in plasma, measured by BCA assay (Pierce), did not differ between groups. Experimental procedures were approved by IACUC.

[0290] Human Studies. Thirteen healthy women between the ages of 20-63, body mass indexes (in kg/m²) between 20.2 and 31.9, were included for this study. Inclusion was contingent on a good, age-typical health status, as ascertained by physical examination and standard clinical laboratory tests such as complete blood count, blood chemistries, fasting glucose, insulin, lipid and liver function tests, liver lipid content and the presence of no known chronic disease including diabetes. Subjects were admitted to the Clinical and Translational Research Center of UNC and placed on a balanced weight maintenance diet for 10 days (Fischer et al., *J. Clin. Nutr.* 85:1275 (2007)). Circulating SOGA and adiponectin were measured from plasma samples collected from

an intravenous catheter following an overnight fast. The race-ethnicity distribution of the participants was white (63%), African American (27%), Asian (6%), and Native American (4%), which reflected the local population characteristics of the Raleigh-Durham-Chapel Hill area. Plasma adiponectin and SOGA were determined by SDS-PAGE using polyclonal antisera against human adiponectin and human SOGA, horseradish peroxidase linked secondary anti-rabbit IgG. Circulating adiponectin and SOGA levels were measured by enhanced chemiluminescence (ECL) signal intensity. Human studies were performed under an IRB approved protocol (CTRC-2645; Study: 07-1158).

[0291] Statistical Analysis. Student's t test was used to identify significant differences when data within groups showed a normal distribution and Wilcoxon-Rank Sum test was used when data did not show a normal distribution. P values less than 0.05 were considered significant.

Example 3

Identification of SOGA by Mass Spectrometry

[0292] Protein extracts from hepatoma cells exposed to adiponectin were digested with trypsin and analyzed by mass spectrometry. Mass spectrometry revealed a peptide, KVLP-SEEDDFLEVNSM (SEQ ID NO:8), encoded by a gene located on chromosome 2 in mice (2411) and chromosome 20 in humans (20q 11). Mouse liver RNA was used to clone the full length 4.7 kb SOGA cDNA (GENBANK ID: H977045). Northern blot analysis, using a probe recognizing the C-terminal end of SOGA, revealed a single dominant 4-5 kb band in the liver. The ORF of the cDNA clone predicts a 161 kDa protein that contains an internal secretory peptide sequence, FKHNFLLLMKLRWFLKRWRQG (SEQ ID NO:9) (FIG. 10). On the basis of computational methods that incorporate signal peptide and cleavage site predictions, SOGA is cleaved between G at the end of the signal peptide and K at the beginning of the peptide identified by mass spectrometry (Emanuelsson et al., *Nat. Protoc.* 2:953 (2007)).

[0293] FIG. 10A is a map showing the location of the conserved ATG16 and Rab5-binding motifs, the secretory signal peptide and the species-specific epitope in the predicted 161 kDa SOGA. The map also shows the predicted domains of the 80 kDa peptide detected *in vitro* and the 25 kDa peptide detected in plasma. FIG. 10B shows the amino acid sequence for murine SOGA (SEQ ID NO:2) showing the location of the Atg16 (232-375) and Rab5-binding (757-886) motifs underlined, the signal peptide (681-702) in bold, the tryptic peptide identified by mass spectrometry (703-718) shaded and the species specific domain (1392-1416) in a box. The position of the internal signal peptide explains why our antibodies, recognizing the species-specific epitope near the C-terminus of SOGA, detect an 80 kDa SOGA peptide rather than the 161 kDa SOGA protein.

Example 4

Function of SOGA in Primary Hepatocytes

[0294] Consistent with the predicted position of the cleavage site, rabbit antisera recognizing the species-specific domain on the C-terminal region of murine SOGA recognized a single 80 kDa protein in isolated hepatocytes (FIG. 11A). FIG. 11A shows a representative SDS-PAGE of primary murine hepatocyte samples showing the knockdown of 80 kDa SOGA as a function of time after exposure to siRNA. siRNA suppression of SOGA caused a dramatic increase in lysosome and late autophagic vacuole number (2.0 ± 0.2 per cell compared to 17.5 ± 2.0 per cell where n=25-30 cells per

group, p<0.0001) as indicated by isolated punctate acidotropic dye staining which provides correlative data on autophagy (FIG. 11B) (Klionsky et al., *Autophagy* 4:151 (2008)). FIG. 11B shows representative purified binucleate hepatocyte cultures transfected with control (left) or SOGA siRNA (right) stained with the lysosome-specific fluorescent dye Lysotracker Red. The hypothesis that SOGA inhibits autophagy is further supported by the reduction of total cell protein content 48 hours after siRNA suppression of SOGA (11.2 ± 0.6 μ g/well compared to 16.3 ± 0.4 μ g/well; n=4 per group; p<0.05). FIG. 2C depicts bar graphs showing the effects of adiponectin and SOGA siRNA on glucose and valine secretion in hepatocyte conditioned media (top and middle) and 80 kDa SOGA measured by densitometry of ECL (enhanced chemiluminescent signal) after SDS-PAGE (bottom). Adiponectin exposure caused a 40% increase of SOGA in primary hepatocytes and a 50% reduction in glucose production (FIG. 11C). siRNA suppression of SOGA blocked the inhibition of glucose production and stimulated valine secretion (FIG. 11C). The secretion of valine, an essential amino acid that cannot be metabolized, due to the absence of branched chain aminotransferase in hepatocytes, also suggests an increase in autophagy. These results support the hypothesis that the elevation of SOGA in response to adiponectin exposure is linked to the inhibition of autophagy.

Example 5

Regulation of SOGA in Primary Hepatocytes and the Correlation of Intracellular and Extracellular Levels of SOGA

[0295] FIG. 11D depicts bar graphs showing the roles of AMPK and PI3K on adiponectin regulation of intracellular and extracellular SOGA levels. Primary hepatocytes were incubated in the presence or absence of 500 μ M AICAR, a stimulator of AMPK, or 10 nM LY294002, a PI3K inhibitor. Bars represent mean values \pm SEM for n=4 per group where “**” indicates a significant difference compared to control (left bar) at p<0.05 by nonparametric Student's t-test. Grey and black bars indicate whether measurements were made in hepatocyte conditioned media or hepatocytes, respectively. The activation of AMPK by AICAR caused a decrease in SOGA that was blocked by adiponectin exposure (FIG. 11D). On the other hand, the inhibition of PI3K by LY294002 caused a decrease in SOGA that was not blocked by adiponectin (FIG. 11D). These observations suggest that adiponectin increases SOGA through the insulin signaling pathway through a mechanism that can be inhibited by AMPK. Consistent with the identification of an internal secretory signal peptide in SOGA, SDS-PAGE analysis revealed that the 80 kDa SOGA fragment is secreted in hepatocyte conditioned media. The reduction of intracellular SOGA by adiponectin and LY294002 was reflected in the levels of SOGA in hepatocyte conditioned media. These results suggested that extracellular levels of SOGA could be used as a biomarker of its intracellular activity.

Example 6

Circulating SOGA in Mice and Humans

[0296] Antisera from 2 different rabbits immunized with two different peptide antigens, 476 and 477, detected a 25 kDa peptide in mouse plasma (FIG. 12A). SDS-PAGE shows the SOGA peptide antigen 476 was detected with immune but not pre-immune sera. The blot exposed to immune sera shows that the signal intensity is proportional to the peptide antigen concentration. FIG. 12B, left panel, shows that mouse-spe-

cific polyclonal antisera 476 detected a 25 kDa protein in mouse plasma but not human plasma. FIG. 12B, right panel, shows that antisera from two different rabbits immunized with two different peptide antigens, 476 and 477, detected a 25 kDa peptide in mouse plasma. Peptide antigens 476 and 477 correspond to overlapping amino acid sequences in the species specific epitope of SOGA. Peptide antigens used to produce rabbit antisera, SAQSLASCFCIRPSRNPIRFISPSKC (SEQ ID NO:7) (antigen 476) and CSAQSLASCFCIRPSRN (SEQ ID NO:6) (antigen 477), were analyzed by mass spectrometry to confirm their amino acid sequence. Rabbit antisera recognizing murine SOGA did not cross-react with any proteins in human plasma. FIG. 12C, top panel, shows a UV absorption plot for plasma proteins generated by HPLC. SDS-PAGE shows that 25 kDa SOGA eluted in fraction 9. For reference, the triglyceride peak (VLDL particle, ~400 kDa) and the cholesterol peak (HDL particle, ~200 kDa) were observed in fractions 1-2 and 5-6, respectively. HPLC analysis confirms that 25 kDa SOGA circulates as a monomer. FIG. 12C, bottom panel, presents SDS-PAGE showing SOGA precipitated out of HPLC fraction 9 in a 40% ammonium sulfate solution. Due to the presence of cysteine residues within the antigenic motif of SOGA, antibody detection of 25 kDa SOGA required the reduction of the sample with dithiothreitol. Based on the predicted sequence of 25 kDa fragment, the intramolecular disulfide bonds between cysteine residues on the carboxy-terminal end of SOGA should generate a fish hook conformation. Two observations indicate that 25 kDa SOGA circulates as a monomer. First, SOGA was detected at 25 kDa when plasma samples were reduced after SDS-PAGE. Second, by size exclusion chromatography of plasma proteins under native conditions, SOGA eluted at 25 kDa (FIG. 12C).

[0297] Recombinant 25 kDa SOGA was produced in *E. coli* and was detectable with the antibodies raised against full length SOGA. FIG. 13A shows a BsrG1 digest of murine 25 kDa SOGA clone in pET-DEST42 GATEWAY vector in 2% agarose. FIG. 13B shows a SDS-PAGE blot of recombinant 25 kDa murine SOGA, either without or with 6xHis tag, produced in IPTG stimulated *E. coli* transformed with the pET-DEST42

[0298] GATEWAY vector. The left panel shows cross reactivity of our murine SOGA antisera with recombinant 25 kDa murine SOGA. The right panel shows a Ponceau red stained blot of total bacterial lysates after SDS-PAGE.

Example 7

Correlation between Circulating Adiponectin and SOGA

[0299] To further validate the link between adiponectin and SOGA in vivo, circulating levels of adiponectin and SOGA were measured in (a) healthy human volunteers, (b) wild-type mice after weight reduction by calorie restriction, and (c) pioglitazone treatment in ob/ob mice, a model of type II diabetes. FIG. 14A shows adiponectin and 25 kDa SOGA levels in human plasma from healthy female volunteers (ages 20-63; n=13). Plasma was collected after an overnight fast. Values represent averages from 2 plasma samples taken 10 minutes apart. A correlation coefficient (R^2) of 0.82 was found between SOGA and adiponectin. The analysis of human plasma from healthy fasting female volunteers (plasma insulin: $7.1 \pm 1.0 \mu\text{U}/\text{ml}$) showed a positive correlation between circulating levels of adiponectin and SOGA ($R^2=0.82$) (FIG. 14A). FIG. 14B shows the effect of ad libitum (AL) versus 30% calorie restricted (CR) feeding on adiponectin, SOGA and glucose in wild-type mice. Bar graphs

show levels of plasma adiponectin (top), 25 kDa SOGA (middle) and glucose (bottom). Calorie restriction, a nutritional intervention that doubled plasma adiponectin, resulted in a 2-fold elevation of circulating SOGA (FIG. 14B). The concentration of plasma glucose in calorie restricted mice compared to ad libitum fed mice was $80 \pm 7 \text{ mg/dl}$ and $131 \pm 10 \text{ mg/dl}$, respectively (FIG. 14B). The complex oligomeric structure, high turnover rate and abundance of circulating adiponectin prevented us from using recombinant adiponectin to study the regulation of SOGA in vivo (Shetty et al., *Trends Pharmacol. Sci.* 30:234 (2009)). Therefore, oral pioglitazone treatment was used to elevate adiponectin in ob/ob mice, an obese model of type II diabetes. FIG. 14C shows the effect of pioglitazone treatment on liver SOGA mRNA and circulating adiponectin, SOGA and glucose in diabetic ob/ob mice. Mice received a daily dose of pioglitazone (TZD) or placebo (CTL) by oral gavage. Bar graphs show the levels of plasma adiponectin (top), liver SOGA mRNA/18S RNA (second), plasma 25 kDa SOGA (third) and plasma glucose (bottom) after 4 days of treatment. Pioglitazone treatment caused a 40% increase of SOGA mRNA in the liver and a 3-fold elevation of circulating adiponectin and SOGA (FIG. 14C). The concentration of plasma glucose was $155 \pm 8 \text{ mg/dl}$ in pioglitazone treated ob/ob mice compared to $450 \pm 18 \text{ mg/dl}$ in untreated ob/ob mice ($p < 0.05$) (FIG. 14C). These results support the hypothesis that adiponectin elevation of SOGA increases insulin sensitivity. Both calorie restriction and pioglitazone treatment have pleiotropic effects beyond the elevation of circulating adiponectin making it difficult to draw any conclusions about the linkage between adiponectin and SOGA. Hence, circulating levels of SOGA between wild-type and adiponectin transgenic mice were compared. FIG. 14D shows circulating levels of adiponectin and SOGA in male adiponectin transgenic mice and their wild type litter mates on a high fat diet. Bars in panels B, C and D represent mean \pm SEM for n=4-5 per group where “**” indicates a significant difference ($p < 0.05$) by nonparametric Student’s t-test. Previous studies have shown that the 3-fold elevation of adiponectin in transgenic mice exerts a protective effect against diabetogenic high fat diet (Combs et al., *Endocrinology* 145:367 (2004); Brooks et al., *J. Biol. Chem.* 282: 35069 (2007)). Consistent with a stimulatory effect of adiponectin, circulating levels of SOGA were higher in adiponectin transgenic mice than their wild type litter mates on a high fat diet (FIG. 14D). These results support the hypothesis that the increase of SOGA in response to adiponectin contributes to the reduction of glucose production in vivo.

Example 8

Correlation between Circulating Insulin and SOGA

[0300] Because adiponectin is an insulin sensitizer and the inhibition of the insulin signaling intermediate PI3K blocked the induction of SOGA in isolated hepatocytes (FIG. 11D), we sought to determine whether there is a correlation between circulating insulin and SOGA during (a) feeding and fasting in humans and (b) insulin withdrawal in NOD mice, a model of type I diabetes. FIG. 15A shows the percent change in circulating levels of SOGA in healthy human volunteers (20-43 years old) measured at 8-11 AM, within 2 hours of feeding or following an overnight (10-12 hour) fast. Bars represent mean values \pm SEM for n=5 and “**” indicates a significant difference at $p < 0.05$ by nonparametric Student’s t-test. Consistent with the theory that insulin stimulates SOGA, a 12-hour fast in healthy human volunteers was associated with a 25% decrease in circulating SOGA (FIG. 15A). The reduc-

tion of SOGA in the fasted state is consistent with the induction of SOGA by insulin and the role of SOGA in the inhibition of autophagy and glucose production. FIG. 15B shows the effect of insulin withdrawal and insulin injection on SOGA and glucose in NOD mice. Circulating levels of 25 kDa SOGA and glucose in NOD mice without diabetes (Group 1), NOD mice with diabetes (Group 2) and NOD mice with diabetes treated by a single injection of insulin 24 hours earlier (Group 3) were measured. Bar graphs show the levels of plasma SOGA (top) and glucose (bottom). Bars show mean \pm SEM for n=5 per group where “*” indicates significantly lower than Groups 1 and 3, “**” indicates significantly greater than Group 2 and “***” indicates significantly greater than Groups 1 and 3. Statistical significance was determined by Student's t-test where p<0.05. A 3-fold reduction of circulating SOGA in hyperglycemic NOD mice, in comparison to euglycemic NOD mice, also suggests that insulin induces SOGA in vivo (FIG. 15B). In support of the theory that the increase of SOGA in response to insulin contributes to the reduction of plasma glucose, the treatment of type I diabetes by insulin injection was associated with a 2-fold induction of SOGA (FIG. 15B).

[0301] The results of this study suggest that the elevation of SOGA in response to adiponectin and insulin can lower liver glucose production through the inhibition of autophagy resulting in a decrease of plasma glucose. The observation that knockdown of SOGA elevated glucose production in primary hepatocytes suggested that SOGA is an inhibitor of glucose production. The elevation of glucose production during the reduction of SOGA was linked to changes in primary hepatocytes that suggested an increase in autophagy such as the reduction in protein content and the elevation of lysosome staining and the secretion of valine, a branched chain amino acid that cannot be synthesized or metabolized in hepatocytes.

[0302] The hypothesis that SOGA may interfere with autophagy is supported by the identification of conserved domains found in Atg16 and Rab5-binding proteins (Longatti et al., *Cell Death Differ.* 16:956 (2009)). Both Atg16 and the Rab5-binding proteins contribute to the early stages of autophagy. Although Atg16 is an essential component of the autophagic machinery, adenoviral overexpression of Atg16 inhibits autophagy in mammalian cells (Matsushita et al., *J. Biol. Chem.* 282:6763 (2007)); Fujita et al., *Mol. Biol. Cell* 19:2092 (2008)). The disruption of autophagy by overexpression of Atg16 provides a paradigm that may explain how elevated SOGA inhibits glucose production. Although the current study focuses on the role of SOGA in the liver, it is important to point out that SOGA is also expressed in the other gluconeogenic organs like the kidney and tissues that are rich sources of gluconeogenic substrates like skeletal and cardiac muscle. The elevation of SOGA in extrahepatic tissues may play a critical role in the reduction of glucose production and the amelioration of glucose homeostasis.

[0303] Intracellular levels of SOGA in isolated hepatocytes were proportional to the levels of SOGA in hepatocyte conditioned media leading us to propose that circulating levels of SOGA can be used as a biomarker of intracellular SOGA levels. This hypothesis was supported by the elevation of liver SOGA mRNA and circulating SOGA in pioglitazone treated ob/ob mice. Our in vitro experiments suggest that the elevation of circulating SOGA indicates a decrease in glucose production. This interpretation is consistent with the elevation of circulating SOGA after calorie restriction, oral pioglitazone, transgenic elevation of adiponectin, feeding and insulin injection. Although glucose production was not measured in the present study, previous reports in mice, rats and humans

show that glucose production is reduced by the elevation of adiponectin in transgenic mice, the implementation of calorie restriction, the treatment of type II diabetes by oral insulin sensitizers and the treatment of type 1 diabetes by insulin (Wahren et al., *Annu. Rev. Nutr.* 27:329 (2007); Combs et al., *J. Clin. Invest.* 108:1875 (2001); Combs et al., *Endocrinology* 145:367 (2004); Barzilai et al., *J. Clin. Invest.* 101:1353 (2998); Miyazaki et al., *J. Clin. Endocrinol. Metab.* 89:4312 (2004)).

[0304] The elevation of SOGA in calorie restricted, pioglitazone and adiponectin transgenic mice supports the hypothesis that adiponectin induces SOGA. The elevation of SOGA in response to adiponectin was not impaired by pharmacologic inhibition of AMPK in isolated hepatocytes suggesting that the induction of SOGA is an insulin sensitizing effect of adiponectin that is mediated independent of AMPK. Adiponectin mediated increases in SOGA were impaired by pharmacologic inhibition of the insulin signaling intermediate PI3K suggesting that the expression of SOGA is regulated by the insulin signaling pathway. The reduction of circulating SOGA by a 12-hour fast in humans or hyperglycemic NOD mice and the elevation of circulating SOGA by insulin injection support the hypothesis that SOGA is induced by the insulin signaling pathway. Adiponectin could increase SOGA through the insulin signaling pathway via APPL1, an adaptor protein that binds to the intracellular domain of the adiponectin receptors and the catalytic subunit of PI3K (Mao et al., *Nat. Cell Biol.* 8:516 (2006); Mitsuuchi et al., *Oncogene* 18:4891 (1999); Yang et al., *J. Biol. Chem.* 278:16820 (2003)).

[0305] Antibodies recognizing the C-terminal region of murine SOGA show that cultured hepatocytes as well as liver samples incubated ex vivo secrete an 80 kDa SOGA fragment rather than a 161 kDa protein predicted by the 4.7 kb cDNA. The size discrepancy is explained by the location of an internal secretory signal peptide, also seen in chicken ovalbumin (Lingappa et al., *Nature* 281:117 (1979)). The presence of repeated LXXXXXL sequences in the amino terminal portion of the SOGA (amino acids 222-250 and 288-314) suggests a potential feedback mechanism through protein-protein interactions of leucine zipper motifs in SOGA and APPL1. The absence of 25 kDa SOGA in hepatocytes and liver conditioned media suggests that proteolytic cleavage of 80 kDa SOGA depends on an extracellular factor that is inactive or absent in vitro. The incubation of mouse hepatocyte conditioned media containing 80 kDa SOGA with endothelial cells (HUVECs) or human plasma did not yield a 25 kDa fragment. Circulating SOGA may play a physiologic role in glucose homeostasis.

[0306] The discovery that circulating levels of adiponectin and SOGA were highly correlated in humans suggests that the measurement of SOGA may be clinically relevant. For example, while TZD drug treatment is almost always effective in the induction of adiponectin, it is only effective in lowering glucose in 70% of type II diabetics (Snitler et al., *Diabetes Care* 27:1365 (2004)). Insulin treatment in type I diabetics is also not completely effective 100% of the time. Based on the results presented here, it would not be surprising if specific cases of poor clinical outcomes were associated with poor induction of SOGA.

[0307] The foregoing is illustrative of the present invention, and is not to be construed as limiting thereof. The invention is defined by the following claims, with equivalents of the claims to be included therein.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 11

<210> SEQ ID NO 1

<211> LENGTH: 4716

<212> TYPE: DNA

<213> ORGANISM: *Mus musculus*

<400> SEQUENCE: 1

agttgggcct ggagctggcg ctgagcagcg acgccgagtc tgccggggc ggcggggcg 60
ggacccgcac cgggcagccg ccccagccag cgcaagtggg gcagcagccct ccggggcctc 120
ccgcctcccc ggatgagccg tcggtgccg catcgctggt gggcagcagc cgcttgccat 180
tcagegcctc gctageccttc tccgacacta ccgaggagat gctggactgt gggccggag 240
gcttggtgcg ggagctggaa gagctgcgtt ccgagaacga ctatctaag gatgagattg 300
aggagctacg ggctgagatg ctggagatgc gggatgtcta catggaggaa gacgtgtatc 360
agctgcagta ccgactgcgt aaggctgacg gccgcagccct ccgcgcgtgcc cagacaggcc 420
aggttcatgg ggaactcatc cgaggtctgg aacaggacgt caaggctct aaggacatct 480
ccatgcggct tcacaaggag ctggaggatgg tggagaagaa gcggatgagg ctggaggagg 540
agaacgaggg gtttcgacag aggetcattt agacagagct ggccaagcag gtgtacaga 600
cggagctgga tcgtcccaga gaggattct tgaagaaaag aggaacccgg tctctgggaa 660
agacagataa gaagcctact gcacaggagg atagtgcaga cctgaagtgc cagctgcatt 720
ttgcaaagga ggagtcggcc ctcatgtgca agaagctcac caagttggct aaggagaacg 780
acagcatgaa ggaggagctg ctcaagtaca gatcgctcta tggggacctg gatgcagccc 840
tgtcggcaga ggagctggcg gatgctccgc actccctgtca gactgagctg aaggtgcacc 900
tgaagctggt ggaggaggag gccaacctgc tgagccggc catagtggag ctggaggatgg 960
agaacccgtgg cctgcgcgac gagatggacg acatgaagga ccacgggggt ggccggggtc 1020
ccgaggccag gctggccttc tcttcctgg gtggtgatgt cggggagacg ctggccagat 1080
tgcggcgcca cctgcgttc gtggaagagg aggctgacgt gctgaggcgc tcctcagctg 1140
agctggagga ccagaacaag ttgctgtca acgagctggc caaataccgc tcggagcagc 1200
agctggacgt gacgcgtgtcg gaggacagct gctccgtct cagcgagccc tcgcaggagg 1260
agctggcagc cgccaaagctg cagatcgccg agctcagccg caaggtaag aagctgcagt 1320
atgagaaccg cgtgccttc tccaaatctgc agcgctgtca cctggcctcc tgccagagca 1380
cacggccat gctggagacg gacgcgtgagg ctggggactc tgccgcgtgc gtgcctgccc 1440
ctctgggtga gacgcgtggag cccacgcgcg cccggctgtc caggggccgt gaagccgagg 1500
cgctggccgg cctacggggag caggccgtt tggtcagcaa ggccatcgac gtccctgggg 1560
ctgatgcca tggcttctca gtcggcctcc gctgtgcct ggacaatgag tgtgctgact 1620
tgcgactgca cgaggccct gacaacacgc agggccccag ggtgcacag ctcattccacg 1680
ccatctgggt ggggcgtgatgttgcac aggagctgaa cgccttcacc cgcaaggccag 1740
atgtggccctt ggggagctct ggcaaggagc agcctgagcc cttccctgtct ctgcctgcct 1800
tgggctccca gggccctgtct aaggagatca tgctgtccaa agacattggc tctgacttcc 1860
agccacactga cttcagagac ctgcgttgatgt gggagccag gatccgagag gcctccgtat 1920

-continued

ccggggactt ggagtccaag cctgacccta gtcggaaacctt caggcccata cgagctgaag	1980
ataacgatcc ttatgcctct gagatcaagg atcttcagct ggtcctggcc gagggcccaa	2040
acagectccg gggcttgcaa gageagctgt cccaggagcg gcagctccgg aaggaggagg	2100
ctgacagctt caaccagaaa atggtccagc tgaaggaaga ccagcagagg gcgtcgat	2160
gacgggagtt tgagctgcag agtctgagcc tccagcggcg actggagcag aagttctgga	2220
gccaagagaa gaacatcctg gtgcaggagt cccagcagtt caagcacaac ttctcgat	2280
tcttcatgaa gtcctgggg ttcctgaagc gctggggca gggcaaggtt ctgcccagcg	2340
aagaggatga cttccctggag gtgaacagca tgaaggaact gtacctgctg atggaggaag	2400
aggagatgaa cgcgcagcac tcggataaca aggccctgcac aggggagagc tggacccaga	2460
acacgcctaa tgagtgcatt aagaccctgg ccgacatgaa ggtcacccctg aaggagctgt	2520
gctggctgct ccaggacgag cgtcggggtc tgactgaact tcagcagcag ttcgaaagg	2580
ccaaggccac ctgggagaca gagcgtgcag agctcaaggg ccacgcctcg cagatggagc	2640
tgaaggctgg gaagggtgcc agttagaggg ccgggcctga ctggaaaggct gcactgcaga	2700
gagagcggga ggagcagcaa cacctctgg cagactccta cagcgcgcgc atggagctga	2760
cgagggcagct gcagctgagc gagegccact ggagccagga gaagctgcag ctgggtggagc	2820
ggctgcagg agaaaagcag caggtggagc agcaggtgaa ggagctgcag aaccgcctca	2880
gtcagttgca gaaggctgcc gagccctggg tcctgaagca ctcagacatg gagaagcaag	2940
acaacagctg gaaagaggca cgaagtgaga agacccatga caaggagggt gtctctgaag	3000
ctgagctcgg gggactggc ttaaagagga ccaaattcgt ctctccatg tctgagtttg	3060
aaagtttgc ctactgtcc ccttacatttgc ctggcgggga tgcccgaaac aagaagctgc	3120
ccaaacggccc tgctttgcc tttgtgagta ctgagccagt ggagcctgag aaagacgcca	3180
aggagaaggc ggggctttcc acccggact gttagccacat tggtagtttgc cctgtcagg	3240
aacctgcagg gagacagatg cagcgcagct acacggctcc agacaagacg ggaatccgag	3300
tctactatag tccgcgcagg gtcggcgcc tgggtgtccc tgtggtccat gacaaggagg	3360
gcaagatcct cattgagcca ggcttcctct tcactaccgc caagccaaag gagtcagccg	3420
aggctgacgg gctggccgag agcttcata gcccgggtt ttgcaatttc tcccgccagc	3480
ggctggatgg agatccggg gccagcacct cgggtccgg acctgcttcc cccgccttgc	3540
atgactttga gatgtgggc aacatgagtg acgacatgaa ggagatcacc aactgcgtgc	3600
ggcaggccat gcgctccggc tctctggaga ggaaggtaaa gaacacatcc agccagacgg	3660
taggcgtggc caccgtgggc acccagacca ttccggacggt cagtgttaggt ctgcagccg	3720
acccaccccg cagcgcctc cacagcaaga gtcggatcc cccgcgtcc tgcgttgtgt	3780
ctgtgcgcag caagcagatc tcttcctccc tggacaagggt ccattctcgat attgagcggc	3840
catgttgctc gcccacgtac ggctcaccca agctccagag acgatcggtt tccaaagctgg	3900
atagcaccaa ggaccgcagc ctgtggaaacc tgcaccaggaa caagcaaaat ggctccgcct	3960
gggcgtcgctc caccaccaca cgggatagcc ctgtactgag gaacatcaat gatgggcctt	4020
ctagcctctt tagtgtgggt gggactctg ggagcaccga gtctgtgtgg aaactgggca	4080
tgtctgaggc cccaaaccata cctgagcctc ccaagttatgg cattgtttagt gagttttcc	4140
ggaacgtgtg tggccggca ccggccccca ctactgcago aggcgaggaa agctgcaga	4200

-continued

```
aaccagagcc ctttcgcca gccagctacc atcaaccga gggtatcc aggatctga 4260
acaagaaggc ggccaaggca ggtggtagcg aagaggtcg acccaccatg ctgtcccagg 4320
tgggaagga tggcatttc ttggatggag atggatcattt gatcctcccc agtgaggatg 4380
ccgtatgtga ctgttagcgcc cagtcaactt cctctgtttt catccggcca tcccgcaaca 4440
ccatccggca ctctcatttc aagtgcaggc tgacaccctt agagtcaggc tggggcgggg 4500
aggagagggc agctcccccag tgagtcctt agcaaccaag cacccaccc aagcagccca 4560
gaccctggag atgaggcaag ggctcggtc ctcagctca gtccatccag gagaaatggc 4620
agctgtgcca ctgccacaga agagcttca cattaaggta aagcaagggtg tcttgctgac 4680
tgctggcag tgacctctga tttccaggaa aagaca 4716
```

<210> SEQ ID NO 2

<211> LENGTH: 1434

<212> TYPE: PRT

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 2

```
Met Leu Asp Cys Gly Pro Gly Gly Leu Val Arg Glu Leu Glu Glu Leu
1 5 10 15
Arg Ser Glu Asn Asp Tyr Leu Lys Asp Glu Ile Glu Glu Leu Arg Ala
20 25 30
Glu Met Leu Glu Met Arg Asp Val Tyr Met Glu Glu Asp Val Tyr Gln
35 40 45
Leu Gln Tyr Arg Leu Arg Lys Ala Glu Arg Arg Ser Leu Arg Ala Ala
50 55 60
Gln Thr Gly Gln Val Asp Gly Glu Leu Ile Arg Gly Leu Glu Gln Asp
65 70 75 80
Val Lys Val Ser Lys Asp Ile Ser Met Arg Leu His Lys Glu Leu Glu
85 90 95
Val Val Glu Lys Lys Arg Met Arg Leu Glu Glu Asn Glu Gly Leu
100 105 110
Arg Gln Arg Leu Ile Glu Thr Glu Leu Ala Lys Gln Val Leu Gln Thr
115 120 125
Glu Leu Asp Arg Pro Arg Glu His Ser Leu Lys Lys Arg Gly Thr Arg
130 135 140
Ser Leu Gly Lys Thr Asp Lys Lys Pro Thr Ala Gln Glu Asp Ser Ala
145 150 155 160
Asp Leu Lys Cys Gln Leu His Phe Ala Lys Glu Glu Ser Ala Leu Met
165 170 175
Cys Lys Lys Leu Thr Lys Leu Ala Lys Glu Asn Asp Ser Met Lys Glu
180 185 190
Glu Leu Leu Lys Tyr Arg Ser Leu Tyr Gly Asp Leu Asp Ala Ala Leu
195 200 205
Ser Ala Glu Glu Leu Ala Asp Ala Pro His Ser Arg Glu Thr Glu Leu
210 215 220
Lys Val His Leu Lys Leu Val Glu Glu Ala Asn Leu Leu Ser Arg
225 230 235 240
Arg Ile Val Glu Leu Glu Val Glu Asn Arg Gly Leu Arg Ala Glu Met
245 250 255
Asp Asp Met Lys Asp His Gly Gly Gly Pro Glu Ala Arg Leu
260 265 270
```

-continued

Ala Phe Ser Ser Leu Gly Gly Glu Cys Gly Glu Ser Leu Ala Glu Leu
 275 280 285

Arg Arg His Leu Gln Phe Val Glu Glu Glu Ala Glu Leu Leu Arg Arg
 290 295 300

Ser Ser Ala Glu Leu Glu Asp Gln Asn Lys Leu Leu Leu Asn Glu Leu
 305 310 315 320

Ala Lys Tyr Arg Ser Glu His Glu Leu Asp Val Thr Leu Ser Glu Asp
 325 330 335

Ser Cys Ser Val Leu Ser Glu Pro Ser Gln Glu Glu Leu Ala Ala Ala
 340 345 350

Lys Leu Gln Ile Gly Glu Leu Ser Gly Lys Val Lys Lys Leu Gln Tyr
 355 360 365

Glu Asn Arg Val Leu Leu Ser Asn Leu Gln Arg Cys Asp Leu Ala Ser
 370 375 380

Cys Gln Ser Thr Arg Pro Met Leu Glu Thr Asp Ala Glu Ala Gly Asp
 385 390 395 400

Ser Ala Gln Cys Val Pro Ala Pro Leu Gly Glu Thr Leu Glu Pro His
 405 410 415

Ala Ala Arg Leu Cys Arg Ala Arg Glu Ala Glu Ala Leu Pro Gly Leu
 420 425 430

Arg Glu Gln Ala Ala Leu Val Ser Lys Ala Ile Asp Val Leu Val Ala
 435 440 445

Asp Ala Asn Gly Phe Ser Val Gly Leu Arg Leu Cys Leu Asp Asn Glu
 450 455 460

Cys Ala Asp Leu Arg Leu His Glu Ala Pro Asp Asn Ser Glu Gly Pro
 465 470 475 480

Arg Asp Ala Lys Leu Ile His Ala Ile Leu Val Arg Leu Ser Val Leu
 485 490 495

Gln Gln Glu Leu Asn Ala Phe Thr Arg Lys Ala Asp Val Ala Leu Gly
 500 505 510

Ser Ser Gly Lys Glu Gln Pro Glu Pro Phe Pro Ala Leu Pro Ala Leu
 515 520 525

Gly Ser Gln Gly Pro Ala Lys Glu Ile Met Leu Ser Lys Asp Leu Gly
 530 535 540

Ser Asp Phe Gln Pro Pro Asp Phe Arg Asp Leu Leu Glu Trp Glu Pro
 545 550 555 560

Arg Ile Arg Glu Ala Phe Arg Thr Gly Asp Leu Glu Ser Lys Pro Asp
 565 570 575

Pro Ser Arg Asn Phe Arg Pro Tyr Arg Ala Glu Asp Asn Asp Ser Tyr
 580 585 590

Ala Ser Glu Ile Lys Asp Leu Gln Leu Val Leu Ala Glu Ala His Asp
 595 600 605

Ser Leu Arg Gly Leu Gln Glu Gln Leu Ser Gln Glu Arg Gln Leu Arg
 610 615 620

Lys Glu Glu Ala Asp Ser Phe Asn Gln Lys Met Val Gln Leu Lys Glu
 625 630 635 640

Asp Gln Gln Arg Ala Leu Leu Arg Arg Glu Phe Glu Leu Gln Ser Leu
 645 650 655

Ser Leu Gln Arg Arg Leu Glu Gln Lys Phe Trp Ser Gln Glu Lys Asn
 660 665 670

-continued

Ile	Leu	Val	Gln	Glu	Ser	Gln	Gln	Phe	Lys	His	Asn	Phe	Leu	Leu	Leu
675															685
Phe	Met	Lys	Leu	Arg	Trp	Phe	Leu	Lys	Arg	Trp	Arg	Gln	Gly	Lys	Val
690															700
Leu	Pro	Ser	Glu	Glu	Asp	Asp	Phe	Leu	Glu	Val	Asn	Ser	Met	Lys	Glu
705															720
Leu	Tyr	Leu	Leu	Met	Glu	Glu	Glu	Met	Asn	Ala	Gln	His	Ser	Asp	
725															735
Asn	Lys	Ala	Cys	Thr	Gly	Glu	Ser	Trp	Thr	Gln	Asn	Thr	Pro	Asn	Glu
740															750
Cys	Ile	Lys	Thr	Leu	Ala	Asp	Met	Lys	Val	Thr	Leu	Lys	Glu	Leu	Cys
755															765
Trp	Leu	Leu	Gln	Asp	Glu	Arg	Arg	Gly	Leu	Thr	Glu	Leu	Gln	Gln	
770															780
Phe	Ala	Lys	Ala	Lys	Ala	Thr	Trp	Glu	Thr	Glu	Arg	Ala	Glu	Leu	Lys
785															800
Gly	His	Ala	Ser	Gln	Met	Glu	Leu	Lys	Ala	Gly	Lys	Gly	Ala	Ser	Glu
805															815
Arg	Pro	Gly	Pro	Asp	Trp	Lys	Ala	Ala	Leu	Gln	Arg	Glu	Arg	Glu	
820															830
Gln	Gln	His	Leu	Leu	Ala	Glu	Ser	Tyr	Ser	Ala	Val	Met	Glu	Leu	Thr
835															845
Arg	Gln	Leu	Gln	Leu	Ser	Glu	Arg	His	Trp	Ser	Gln	Glu	Lys	Leu	Gln
850															860
Leu	Val	Glu	Arg	Leu	Gln	Gly	Glu	Lys	Gln	Gln	Val	Glu	Gln	Gln	Val
865															880
Lys	Glu	Leu	Gln	Asn	Arg	Leu	Ser	Gln	Leu	Gln	Lys	Ala	Ala	Glu	Pro
885															895
Trp	Val	Leu	Lys	His	Ser	Asp	Met	Glu	Lys	Gln	Asp	Asn	Ser	Trp	Lys
900															910
Glu	Ala	Arg	Ser	Glu	Lys	Thr	His	Asp	Lys	Glu	Gly	Val	Ser	Glu	Ala
915															925
Glu	Leu	Gly	Gly	Thr	Gly	Leu	Lys	Arg	Thr	Lys	Ser	Val	Ser	Ser	Met
930															940
Ser	Glu	Phe	Glu	Ser	Leu	Leu	Asp	Cys	Ser	Pro	Tyr	Leu	Ala	Gly	Gly
945															960
Asp	Ala	Arg	Asn	Lys	Lys	Leu	Pro	Asn	Gly	Pro	Ala	Phe	Ala	Phe	Val
965															975
Ser	Thr	Glu	Pro	Val	Glu	Pro	Glu	Lys	Asp	Ala	Lys	Glu	Lys	Ala	Gly
980															990
Leu	Ser	Thr	Arg	Asp	Cys	Ser	His	Ile	Gly	Ser	Leu	Ala	Cys	Gln	Glu
995															1005
Pro	Ala	Gly	Arg	Gln	Met	Gln	Arg	Ser	Tyr	Thr	Ala	Pro	Asp	Lys	
1010															1020
Thr	Gly	Ile	Arg	Val	Tyr	Tyr	Ser	Pro	Pro	Val	Ala	Arg	Arg	Leu	
1025															1035
Gly	Val	Pro	Val	Val	His	Asp	Lys	Glu	Gly	Lys	Ile	Leu	Ile	Glu	
1040															1050
Pro	Gly	Phe	Leu	Phe	Thr	Thr	Ala	Lys	Pro	Lys	Glu	Ser	Ala	Glu	
1055															1065
Ala	Asp	Gly	Leu	Ala	Glu	Ser	Ser	Tyr	Ser	Arg	Trp	Leu	Cys	Asn	

-continued

1070	1075	1080
Phe Ser Arg Gln Arg Leu Asp Gly Gly Ser Gly Ala Ser Thr Ser		
1085	1090	1095
Gly Ser Gly Pro Ala Phe Pro Ala Leu His Asp Phe Glu Met Ser		
1100	1105	1110
Gly Asn Met Ser Asp Asp Met Lys Glu Ile Thr Asn Cys Val Arg		
1115	1120	1125
Gln Ala Met Arg Ser Gly Ser Leu Glu Arg Lys Val Lys Asn Thr		
1130	1135	1140
Ser Ser Gln Thr Val Gly Val Ala Thr Val Gly Thr Gln Thr Ile		
1145	1150	1155
Arg Thr Val Ser Val Gly Leu Gln Thr Asp Pro Pro Arg Ser Ser		
1160	1165	1170
Leu His Ser Lys Ser Trp Ser Pro Arg Ser Ser Ser Leu Val Ser		
1175	1180	1185
Val Arg Ser Lys Gln Ile Ser Ser Ser Leu Asp Lys Val His Ser		
1190	1195	1200
Arg Ile Glu Arg Pro Cys Cys Ser Pro Lys Tyr Gly Ser Pro Lys		
1205	1210	1215
Leu Gln Arg Arg Ser Val Ser Lys Leu Asp Ser Thr Lys Asp Arg		
1220	1225	1230
Ser Leu Trp Asn Leu His Gln Gly Lys Gln Asn Gly Ser Ala Trp		
1235	1240	1245
Ala Arg Ser Thr Thr Thr Arg Asp Ser Pro Val Leu Arg Asn Ile		
1250	1255	1260
Asn Asp Gly Leu Ser Ser Leu Phe Ser Val Val Glu His Ser Gly		
1265	1270	1275
Ser Thr Glu Ser Val Trp Lys Leu Gly Met Ser Glu Ala Arg Thr		
1280	1285	1290
Lys Pro Glu Pro Pro Lys Tyr Gly Ile Val Gln Glu Phe Phe Arg		
1295	1300	1305
Asn Val Cys Gly Arg Ala Pro Ser Pro Thr Thr Ala Ala Gly Glu		
1310	1315	1320
Glu Ser Cys Lys Lys Pro Glu Pro Leu Ser Pro Ala Ser Tyr His		
1325	1330	1335
Gln Pro Glu Gly Val Ser Arg Ile Leu Asn Lys Lys Ala Ala Lys		
1340	1345	1350
Ala Gly Gly Ser Glu Glu Val Arg Pro Thr Met Leu Ser Gln Val		
1355	1360	1365
Gly Lys Asp Gly Ile Leu Arg Asp Gly Asp Gly Ser Leu Ile Leu		
1370	1375	1380
Pro Ser Glu Asp Ala Val Cys Asp Cys Ser Ala Gln Ser Leu Ala		
1385	1390	1395
Ser Cys Phe Ile Arg Pro Ser Arg Asn Thr Ile Arg His Ser Pro		
1400	1405	1410
Ser Lys Cys Arg Leu His Pro Ser Glu Ser Gly Trp Gly Gly Glu		
1415	1420	1425
Glu Arg Ala Ala Pro Gln		
1430		

-continued

```

<211> LENGTH: 4862
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

cgctgagcag cgacgcgcag tccgcggccg ggggcgcgcg ggggggtccgt acggggcagc      60
cggcccgagcc cgcgcgcctcc ggcgcgcgcgc ccccgccgcgc gcccgcctcc ccggacgcgc      120
cgtcggtggc cgcgtcgctcg gtgggcgcga gccgcgttgcgc gtcagcgcc tcgttgccct      180
tctccgaccc caccgaggag atgcgtggact gccccccca gggcttggtg cgggagctgg      240
aggagctgcg ctccggagaac gactatctca aggacgagat tgaggagctg cgggcccggaga      300
tgctggagat gccccggacgtc tatatggagg aggaegtgtta tcagctgcag taccggctgc      360
gcaaaaggccga ggcgcgcagt ctccgtgcgc cccagacccgg ccaggtggac ggcgagctta      420
tccgtggctct ggagcaggat gtcaagggtct ctaaggacat ctccatgcgc ctgcataagg      480
agctcgaggt ggtggagaag aaacggggcgc ggctggagga ggagaacgaa gagttcgtc      540
agcggctcat cgagactgag ctggctaagc aggtgtcgca gacggagctg gagcggacccga      600
gagagcattc ttgaagaaaa agaggaaccc gtcctctggg gaaggccgat aagaagactt      660
tggtgcagga ggacagtgcgac gacctgaagt gccagttgcgac ctttgcggaa gaggagtcg      720
ccctcatgtt caagaagctc actaagctt ccaaggagaa tgacagcatg aaggaggagc      780
tgctgaagta cccgcgcgtc tatggggacc tggacagcgc gctgtcgcc gaggagctgg      840
ccgatgcccc ccactcgccgg gagaccggc tgaaggtgcg cctgaagctg gtggaggagg      900
aagccaacctt gctgagccgc cgcatcggtt agctggaggt ggagaacccga ggcctgcggg      960
ctgagatggc cgacatgaag gatcatggag gtggctgtgg gggtccttag gcacgcctgg      1020
ccttcctccgc gctgggtggc ggagagtgcg gggagagctt ggcagagctg cggegacacc      1080
tgcaaggatgtt cgaagaggag gccagactgcgac tggacagcgcctt ctctgcggag ctcqaggacc      1140
agaacaagct gctgctgaac gagctggcca agttccgcgc ggagcacggag ctggacgtgg      1200
cgctgtcgga ggacagttgt tctgtgcgtca gcaaccccttca acaggaggag ctggcggccg      1260
ccaagctgca gatcgccgcgag ctcagcgccgca aggtcaagaa gctgcagttac gagaaccgc      1320
tgctcctctc caacccctccag cgctgtgcacc tcgcctccgt ccagagttac cgcccatgc      1380
tggagacggc cccgcggccgc gggactctg cccagttgtt gcctgctccc ctggcggaga      1440
cacacggatc ccatcgccgcgac cgactctgcgac gagccaggaa ggccggagggtt ctgcctggc      1500
tgagagagca ggccggccctg gtcagtaagg ccatecgatgtt cctgggtggct gatgccaatg      1560
gtttcacggc tggccctccgg ctgtgtctgg acaacggatgt tgctgacttc cggctgcgtt      1620
aggccccccga caacacggcggag ggccccccagg acaccaagctt catccatgcgc atcctggc      1680
gcctgagcgt gctgcagcgcgag gagctgtatgc ctttcacgcgc gaaaggccatgtt gcaatgc      1740
gggtgtctgtt caagggaaacag caggagttctt tctcatcaactt gcccccccttgc ggctcccgagg      1800
ggctctctaa ggagattttt ctggccaaag accttgcgc gacttttcgatcc ccacctgact      1860
tcaggacccctt gccggaaatgg gagccggaggat tccggaggatc tttccgcactt ggtgacttgg      1920
actctaagcc cggcccccggc cggagcttca ggccttaccgc agctgaagac aatgattccct      1980
atgcctctgtt gatcaaggatgtt ctgcagctgg tgcgtggctgtca gggccacggac agcctccggg      2040
gcttgcaaga gcaatgcgttcc caggagccggc agctacgaaa ggaggaggcc gacaatttca      2100

```

-continued

accagaaaaat ggtccagctg aaggaggacc agcagagggc gctcctgagg cgggagttt	2160
agctgcagag tctgagccctc cageggaggc tggagcagaa attctggac caggagaaga	2220
acatgttgtt gcaggagtcc cagecaattca agcacaactt cctgtgctc ttcatgaagc	2280
tcaggtggtt cctcaagcgc tggeggcagg gcaaggttt gcccagegaa gggatgact	2340
tcctcgaggt gaacagcatg aaggagctgt acttgctgtat ggaggaagag gagataaacg	2400
ctcagcattc tgataacaag gcctgcacgg gggacagctg gacccagaac acgccccatg	2460
agtacatcaa gacactggcc gacatgaagg tgacgctgaa ggagctgtgc tggctgctcc	2520
gggatgaaacg ccgtggctcg acggagcttc agcaacagtt tgccaaggcc aaggctacct	2580
gggagacaga gcccccccgag ctcaaggccc atacctccc gatggagctg aagacaggg	2640
agggggcccg ggagcgggca gggcccgact ggaaggcago cctacagcgg gagegtgagg	2700
agcagcagca cctcctagct gagtcctaca ggcgtgtcat ggagctgact cggcagctgc	2760
agatcagtga ggcactggaa agctgcagct ggtggagcgg ctgcagggtg	2820
agaagcagca ggtggagcag caggtgaagg agctgcagaa ccgcctaagc cagctgcaga	2880
aggctgccga cccctgggtc ctgaagcaact cggagctgga gaagcaggac aacagctgga	2940
aggagacacg cagtgagaag atccacgaca aggaggctgt ttccgaaggat gagettggag	3000
gaaatggttt aaagagaacc aaatctgttt cttccatgtc tgagttgaa agtttgcctg	3060
actgttcccc ttaccttgct ggcggagatg cccggggcaaa gaagctgcct aacaaccctg	3120
cctttggctt tgtgagctcc gagccaggaa atccagagaa agacaccaag gagaagcctg	3180
ggctctcgcc gagggactgc aaccacctgg gtgcctggc ctgccaggac ccccccaggaa	3240
ggcagatgca ggcgcagctac acggctcctg acaagacggg catccgagtc tactatagtc	3300
ccccgggtggc cggcgccctc ggagtccctg tggttcatga caaagaggc aagatcatta	3360
tgcagccccg cttcccttcc accacagccaa agcccaaaga qtcggccggag gctgtatggc	3420
tggctgagag ctctatggc cggtggctct gcaacttctc acggcagcgc ctggacggag	3480
gctcagcggg cagccccctcg gccccccggc ctggcttccc agcggccctg catgactttg	3540
agatgtcagg caacatgagt gatgacatga aggagatcac caactgtgtc cgccaggc	3600
tgcgcgtccgg ctcactggag aggaaagtga agagcacato cagccagacg gtggccctgg	3660
ccagtgtggg cacacagacc atccgcacgg tcagegtggg cctgcagacc gacccacccc	3720
gcagcagcct ccatggcaag gcctggtcac cccgcagctc ttgcgtcggt tctgtgcgc	3780
gcaagcagat ctctcttcc ctggacaagg tccattcgcc catcgagcgg ccctgctgt	3840
cccccaagta tggctcacca aagctccaga ggcggctgtgt gtccaaagctg gacagcaga	3900
aggaccgcag cctgtggAAC ctgcaccagg gcaagcagaa cggctggcc tggcccccgt	3960
ccaccaccac gccccggacgc cctgtattga gaaacatcaa cgatggactc tccagcctct	4020
tcagtggtt ggagcactca gggagcagg agtctgtctg gaaacttaggc atgtctgaga	4080
cgcggggccaa gcccggccct cccaaagtacg gcattgtgc ggaattttc cgtaatgtgt	4140
gtggccgggc acccgagcccc acctcatcg caggagagga gggcaccaag aagccagacg	4200
ccctctcccc agccagctac catcagccag aggggtggc caggatcctg aacaagaagg	4260
cagccaaagtt gggcagcagt gaggaggtca gactcaccat gtcctccag gtggggaaagg	4320
atggtgttccct cccggacgga gatggagecg tggtccttc caatgaggac gtcgtttgtg	4380

-continued

actgttagtac	ccagtctctc	acctcctgct	tcggccgate	gtccccgtct	gcacatccgcc	4440
actctcccttc	caagtgcagg	ctgcaccctt	cagagtccag	ctgggggtggg	gaggagaggg	4500
cactcccccc	cagcgagtga	cagagcagcc	aagctccccc	cctcaaccag	cccagcccc	4560
ggatagcaga	agggAACCAAG	cagagacgag	acgaggttag	gcgaggggct	gtgtcctcag	4620
cattgcctgg	ccctggaggg	acagcagtga	tgccactgcc	agaatgcagc	tttcacatca	4680
aggtaaagcc	gggtctcctg	ctggcccttg	ggtgttgagc	ttcgacttcc	caggggaaagg	4740
cagtgagtgg	gagagagacc	aaacctgggc	ttccaaagca	tccactgaga	gatctgtcaa	4800
gagccgatcc	ctgggtccata	agagagagcc	ttgcctggtt	ctgcccattgc	caccctcttg	4860
ga						4862

<210> SEQ ID NO 4

<211> LENGTH: 1439

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

Met	Leu	Asp	Cys	Gly	Pro	Ser	Gly	Leu	Val	Arg	Glu	Leu	Glu	Leu
1								10				15		

Arg	Ser	Glu	Asn	Asp	Tyr	Leu	Lys	Asp	Glu	Ile	Glu	Glu	Leu	Arg	Ala
						20		25			30				

Glu	Met	Leu	Glu	Met	Arg	Asp	Val	Tyr	Met	Glu	Glu	Asp	Val	Tyr	Gln
	35						40			45					

Leu	Gln	Tyr	Arg	Leu	Arg	Lys	Ala	Glu	Arg	Arg	Ser	Leu	Arg	Ala	Ala
	50					55			60						

Gln	Thr	Gly	Gln	Val	Asp	Gly	Glu	Leu	Ile	Arg	Gly	Leu	Glu	Gln	Asp
	65					70			75			80			

Val	Lys	Val	Ser	Lys	Asp	Ile	Ser	Met	Arg	Leu	His	Lys	Glu	Leu	Glu
						85		90		95					

Val	Val	Glu	Lys	Lys	Arg	Ala	Arg	Leu	Glu	Glu	Asn	Glu	Glu	Leu	
		100				105			110						

Arg	Gln	Arg	Leu	Ile	Glu	Thr	Glu	Leu	Ala	Lys	Gln	Val	Leu	Gln	Thr
	115					120				125					

Glu	Leu	Glu	Arg	Pro	Arg	Glu	His	Ser	Leu	Lys	Lys	Arg	Gly	Thr	Arg
	130					135			140						

Ser	Leu	Gly	Lys	Ala	Asp	Lys	Lys	Thr	Leu	Val	Gln	Glu	Asp	Ser	Ala
	145					150			155			160			

Asp	Leu	Lys	Cys	Gln	Leu	His	Phe	Ala	Lys	Glu	Glu	Ser	Ala	Leu	Met
	165					170			175						

Cys	Lys	Lys	Leu	Thr	Lys	Leu	Ala	Lys	Glu	Asn	Asp	Ser	Met	Lys	Glu
	180					185			190						

Glu	Leu	Leu	Lys	Tyr	Arg	Ser	Leu	Tyr	Gly	Asp	Leu	Asp	Ser	Ala	Leu
	195					200			205						

Ser	Ala	Glu	Glu	Leu	Ala	Asp	Ala	Pro	His	Ser	Arg	Glu	Thr	Glu	Leu
	210					215			220						

Lys	Val	His	Leu	Lys	Leu	Val	Glu	Glu	Ala	Asn	Leu	Leu	Ser	Arg	
	225					230			235			240			

Arg	Ile	Val	Glu	Leu	Glu	Val	Glu	Asn	Arg	Gly	Leu	Arg	Ala	Glu	Met
	245					250			255						

Asp	Asp	Met	Lys	Asp	His	Gly	Gly	Cys	Gly	Gly	Pro	Glu	Ala	Arg	
	260					265			270						

-continued

Leu Ala Phe Ser Ala Leu Gly Gly Gly Glu Cys Gly Glu Ser Leu Ala
 275 280 285

 Glu Leu Arg Arg His Leu Gln Phe Val Glu Glu Glu Ala Glu Leu Leu
 290 295 300

 Arg Arg Ser Ser Ala Glu Leu Glu Asp Gln Asn Lys Leu Leu Leu Asn
 305 310 315 320

 Glu Leu Ala Lys Phe Arg Ser Glu His Glu Leu Asp Val Ala Leu Ser
 325 330 335

 Glu Asp Ser Cys Ser Val Leu Ser Glu Pro Ser Gln Glu Glu Leu Ala
 340 345 350

 Ala Ala Lys Leu Gln Ile Gly Glu Leu Ser Gly Lys Val Lys Lys Leu
 355 360 365

 Gln Tyr Glu Asn Arg Val Leu Leu Ser Asn Leu Gln Arg Cys Asp Leu
 370 375 380

 Ala Ser Cys Gln Ser Thr Arg Pro Met Leu Glu Thr Asp Ala Glu Ala
 385 390 395 400

 Gly Asp Ser Ala Gln Cys Val Pro Ala Pro Leu Gly Glu Thr His Glu
 405 410 415

 Ser His Ala Val Arg Leu Cys Arg Ala Arg Glu Ala Glu Val Leu Pro
 420 425 430

 Gly Leu Arg Glu Gln Ala Ala Leu Val Ser Lys Ala Ile Asp Val Leu
 435 440 445

 Val Ala Asp Ala Asn Gly Phe Thr Ala Gly Leu Arg Leu Cys Leu Asp
 450 455 460

 Asn Glu Cys Ala Asp Phe Arg Leu His Glu Ala Pro Asp Asn Ser Glu
 465 470 475 480

 Gly Pro Arg Asp Thr Lys Leu Ile His Ala Ile Leu Val Arg Leu Ser
 485 490 495

 Val Leu Gln Gln Glu Leu Asn Ala Phe Thr Arg Lys Ala Asp Ala Val
 500 505 510

 Leu Gly Cys Ser Val Lys Glu Gln Gln Glu Ser Phe Ser Ser Leu Pro
 515 520 525

 Pro Leu Gly Ser Gln Gly Leu Ser Lys Glu Ile Leu Leu Ala Lys Asp
 530 535 540

 Leu Gly Ser Asp Phe Gln Pro Pro Asp Phe Arg Asp Leu Pro Glu Trp
 545 550 555 560

 Glu Pro Arg Ile Arg Glu Ala Phe Arg Thr Gly Asp Leu Asp Ser Lys
 565 570 575

 Pro Asp Pro Ser Arg Ser Phe Arg Pro Tyr Arg Ala Glu Asp Asn Asp
 580 585 590

 Ser Tyr Ala Ser Glu Ile Lys Glu Leu Gln Leu Val Leu Ala Glu Ala
 595 600 605

 His Asp Ser Leu Arg Gly Leu Gln Glu Gln Leu Ser Gln Glu Arg Gln
 610 615 620

 Leu Arg Lys Glu Glu Ala Asp Asn Phe Asn Gln Lys Met Val Gln Leu
 625 630 635 640

 Lys Glu Asp Gln Gln Arg Ala Leu Leu Arg Arg Glu Phe Glu Leu Gln
 645 650 655

 Ser Leu Ser Leu Gln Arg Arg Leu Glu Gln Lys Phe Trp Ser Gln Glu
 660 665 670

-continued

Lys	Asn	Met	Leu	Val	Gln	Glu	Ser	Gln	Gln	Phe	Lys	His	Asn	Phe	Leu
675															685
Leu	Leu	Phe	Met	Lys	Leu	Arg	Trp	Phe	Leu	Lys	Arg	Trp	Arg	Gln	Gly
690															700
Lys	Val	Leu	Pro	Ser	Glu	Gly	Asp	Asp	Phe	Leu	Glu	Val	Asn	Ser	Met
705															720
Lys	Glu	Leu	Tyr	Leu	Leu	Met	Glu	Glu	Glu	Ile	Asn	Ala	Gln	His	
725															735
Ser	Asp	Asn	Lys	Ala	Cys	Thr	Gly	Asp	Ser	Trp	Thr	Gln	Asn	Thr	Pro
740															750
Asn	Glu	Tyr	Ile	Lys	Thr	Leu	Ala	Asp	Met	Lys	Val	Thr	Leu	Lys	Glu
755															765
Leu	Cys	Trp	Leu	Leu	Arg	Asp	Glu	Arg	Arg	Gly	Leu	Thr	Glu	Leu	Gln
770															780
Gln	Gln	Phe	Ala	Lys	Ala	Lys	Ala	Thr	Trp	Glu	Thr	Glu	Arg	Ala	Glu
785															800
Leu	Lys	Gly	His	Thr	Ser	Gln	Met	Glu	Leu	Lys	Thr	Gly	Lys	Gly	Ala
805															815
Gly	Glu	Arg	Ala	Gly	Pro	Asp	Trp	Lys	Ala	Ala	Leu	Gln	Arg	Glu	Arg
820															830
Glu	Glu	Gln	His	Leu	Leu	Ala	Glu	Ser	Tyr	Ser	Ala	Val	Met	Glu	
835															845
Leu	Thr	Arg	Gln	Leu	Gln	Ile	Ser	Glu	Arg	Asn	Trp	Ser	Gln	Glu	Lys
850															860
Leu	Gln	Leu	Val	Glu	Arg	Leu	Gln	Gly	Glu	Lys	Gln	Gln	Val	Glu	Gln
865															880
Gln	Val	Lys	Glu	Leu	Gln	Asn	Arg	Leu	Ser	Gln	Leu	Gln	Lys	Ala	Ala
885															895
Asp	Pro	Trp	Val	Leu	Lys	His	Ser	Glu	Leu	Glu	Lys	Gln	Asp	Asn	Ser
900															910
Trp	Lys	Glu	Thr	Arg	Ser	Glu	Lys	Ile	His	Asp	Lys	Glu	Ala	Val	Ser
915															925
Glu	Val	Glu	Leu	Gly	Gly	Asn	Gly	Leu	Lys	Arg	Thr	Lys	Ser	Val	Ser
930															940
Ser	Met	Ser	Glu	Phe	Glu	Ser	Leu	Leu	Asp	Cys	Ser	Pro	Tyr	Leu	Ala
945															960
Gly	Gly	Asp	Ala	Arg	Gly	Lys	Lys	Leu	Pro	Asn	Asn	Pro	Ala	Phe	Gly
965															975
Phe	Val	Ser	Ser	Glu	Pro	Gly	Asp	Pro	Glu	Lys	Asp	Thr	Lys	Glu	Lys
980															990
Pro	Gly	Leu	Ser	Ser	Arg	Asp	Cys	Asn	His	Leu	Gly	Ala	Leu	Ala	Cys
995															1005
Gln	Asp	Pro	Pro	Gly	Arg	Gln	Met	Gln	Arg	Ser	Tyr	Thr	Ala	Pro	
1010															1020
Asp	Lys	Thr	Gly	Ile	Arg	Val	Tyr	Tyr	Ser	Pro	Pro	Val	Ala	Arg	
1025															1035
Arg	Leu	Gly	Val	Pro	Val	Val	His	Asp	Lys	Glu	Gly	Lys	Ile	Ile	
1040															1050
Ile	Glu	Pro	Gly	Phe	Leu	Phe	Thr	Thr	Ala	Lys	Pro	Lys	Glu	Ser	
1055															1065
Ala	Glu	Ala	Asp	Gly	Leu	Ala	Glu	Ser	Ser	Tyr	Gly	Arg	Trp	Leu	

-continued

1070	1075	1080
Cys Asn Phe Ser Arg Gln Arg	Leu Asp Gly Gly Ser	Ala Gly Ser
1085	1090	1095
Pro Ser Ala Ala Gly Pro Gly	Phe Pro Ala Ala Leu	His Asp Phe
1100	1105	1110
Glu Met Ser Gly Asn Met Ser	Asp Asp Met Lys Glu	Ile Thr Asn
1115	1120	1125
Cys Val Arg Gln Ala Met Arg	Ser Gly Ser Leu Glu	Arg Lys Val
1130	1135	1140
Lys Ser Thr Ser Ser Gln Thr	Val Gly Leu Ala Ser	Val Gly Thr
1145	1150	1155
Gln Thr Ile Arg Thr Val Ser	Val Gly Leu Gln Thr	Asp Pro Pro
1160	1165	1170
Arg Ser Ser Leu His Gly Lys	Ala Trp Ser Pro Arg	Ser Ser Ser
1175	1180	1185
Leu Val Ser Val Arg Ser Lys	Gln Ile Ser Ser Ser	Leu Asp Lys
1190	1195	1200
Val His Ser Arg Ile Glu Arg	Pro Cys Cys Ser Pro	Lys Tyr Gly
1205	1210	1215
Ser Pro Lys Leu Gln Arg Arg	Ser Val Ser Lys Leu	Asp Ser Ser
1220	1225	1230
Lys Asp Arg Ser Leu Trp Asn	Leu His Gln Gly Lys	Gln Asn Gly
1235	1240	1245
Ser Ala Trp Ala Arg Ser Thr	Thr Thr Arg Asp Ser	Pro Val Leu
1250	1255	1260
Arg Asn Ile Asn Asp Gly Leu	Ser Ser Leu Phe Ser	Val Val Glu
1265	1270	1275
His Ser Gly Ser Thr Glu Ser	Val Trp Lys Leu Gly	Met Ser Glu
1280	1285	1290
Thr Arg Ala Lys Pro Glu Pro	Pro Lys Tyr Gly Ile	Val Gln Glu
1295	1300	1305
Phe Phe Arg Asn Val Cys Gly	Arg Ala Pro Ser Pro	Thr Ser Ser
1310	1315	1320
Ala Gly Glu Glu Gly Thr Lys	Lys Pro Glu Pro Leu	Ser Pro Ala
1325	1330	1335
Ser Tyr His Gln Pro Glu Gly	Val Ala Arg Ile Leu	Asn Lys Lys
1340	1345	1350
Ala Ala Lys Leu Gly Ser Ser	Glu Glu Val Arg Leu	Thr Met Leu
1355	1360	1365
Pro Gln Val Gly Lys Asp Gly	Val Leu Arg Asp Gly	Asp Gly Ala
1370	1375	1380
Val Val Leu Pro Asn Glu Asp	Ala Val Cys Asp Cys	Ser Thr Gln
1385	1390	1395
Ser Leu Thr Ser Cys Phe Ala	Arg Ser Ser Arg Ser	Ala Ile Arg
1400	1405	1410
His Ser Pro Ser Lys Cys Arg	Leu His Pro Ser Glu	Ser Ser Trp
1415	1420	1425
Gly Gly Glu Glu Arg Ala Leu	Pro Pro Ser Glu	
1430	1435	

-continued

<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic human-specific peptide antigen

<400> SEQUENCE: 5

Ser Thr Gln Ser Leu Thr Ser Phe Ala Arg Ser Ser Arg Ser Ala Ile
1 5 10 15

Arg His Ser Pro Ser Lys Cys
20

<210> SEQ ID NO 6
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic murine-specific peptide antigen

<400> SEQUENCE: 6

Cys Ser Ala Gln Ser Leu Ala Ser Cys Phe Ile Arg Pro Ser Arg Asn
1 5 10 15

<210> SEQ ID NO 7
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic overlapping murine-specific peptide
antigens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (8)..(8)
<223> OTHER INFORMATION: Xaa is acemidomethyl cysteine

<400> SEQUENCE: 7

Ser Ala Gln Ser Leu Ala Ser Xaa Phe Ile Arg Pro Ser Arg Asn Pro
1 5 10 15

Ile Arg His Ser Pro Ser Lys Cys
20

<210> SEQ ID NO 8
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: SOGA peptide fragment sequence

<400> SEQUENCE: 8

Lys Val Leu Pro Ser Glu Glu Asp Asp Phe Leu Glu Val Asn Ser Met
1 5 10 15

<210> SEQ ID NO 9
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 9

Phe Lys His Asn Phe Leu Leu Phe Met Lys Leu Arg Trp Phe Leu
1 5 10 15

Lys Arg Trp Arg Gln Gly
20

-continued

```

<210> SEQ ID NO 10
<211> LENGTH: 60359
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 10

agttgggcct ggagctggcg ctgagcagcg acggcgagtc tgccggggc ggccgggg 60
ggacccgcac cgggcgcggc ccccagccag cgcaagtggg gcaggcgcct ccggggcctc 120
ccgcctcccc ggatgagccg tcggtgccg catcgctggt gggcagcgc cgcttgccac 180
tcagegcctc gctagccttc tccgacactca ccgaggagat gctggactgt gggccggag 240
gcttggtgcg ggagctggaa gagctgcgtt ccgagaacga ctatctaag gtggggacct 300
ggggggatg ggggaggcgc agggcgaaa gaggactggt catccggct caagagtcca 360
gggcgcgtgac caagcaggaa gagtgccca ctttcttc ccactttcc ggtgaggagg 420
attatggata agacacccgag taaaggaccc caatgtgcac ttcccaccc gggtaaaaac 480
tgttttagag gacacgttct gatgggttag tgccatgtt cggccaggc tgaccctccc 540
ccccctctaac ctgcccagga ccccggtccc agatcagcag gcttctcggt ctctctgttc 600
tggaaaccaat gtccgtgtgt taggctgcca gggaaaccgt gctatcccgt ggcaaggaga 660
ggcagaacac catcctaagt ctttacccag ggcttcccg gcaggcctgt cctcgccctgc 720
ctttgggc gcatccagtt taaaaggcgtg acatctggaa acatcccgtt ggctcgagtc 780
cactccgcg cagtctaacg gtccacacgc ctgggaggat gatgtcggt gagaactgg 840
tcagaggcca ctccctcaat cccaggatat tgtcttctgt cccctctgag agaggtttgg 900
gcagacttgg cccgtgtcc ctgtggaaag agggaaactgg tccctgtgtt taggggtatt 960
ccctctgtgt gtgggcctgg agagctgtct gggagaagca atcgctgggt actgggcacc 1020
taggtgcagc ccatgctctg gctgcctttt ctttgacatg gtggcagctt tggggacagt 1080
cctgccttgg cttcttgacc ttgggtgtgg aagctggat ctccacggc tgcaactgc 1140
ctgctggaaag aggctacagt attgtctgca cttgacagtt gatggcctgg cacaacaccc 1200
agagtccttgg ggcaggaggc cattgagttt acagcactct aatgacatgt ctgcctgtt 1260
gcacttaact gcctttgtct tctactgtcg ggggagttt ctttaggtat ttgggtgttc 1320
cctcttgctt tcccttgggt cctctatcca gcatgttctc aggaaactcc tgcactactgt 1380
aggacacagg tctatctgac aaaggattat ttaccacctg tttgttagctt gggccctccc 1440
caaaccaggc ctccatcccccc ttttttttc agttaaatgg gagttaagt cttcacaaag 1500
gtgttactca ctgttgcacat ttgggagctg gggttcagtt agagggcgaat gtcttctgaa 1560
gtcagtggtt agagtgaggc ccctgaggat ggagtcagtc ttccgagatg ctccctgtgg 1620
cttcctgaat tcccacccctg agctactcca gacctggacc cgagcaccag atgacacaaa 1680
ctcacatgttc ttgggtttggc ttcttcctgc ccagatccccggcccttcc tggctgttttt 1740
ggaagtggc agacagagtc cctaaactcca tctgaagctt cccatgagta tggatgggg 1800
aaagtggcc caggaggctg ggggtcagttt ccaggttctc cttagagacca agactgggtt 1860
gagttccagt tccaggccca gactgagctt cactttctgc caaccattca gttgagtttt 1920
tggcggaaag ctccaggccc ctgttggatc tggatgttcc tggctgtgtt atcttggca 1980
cattacttac catgtttggaa cctgagttttt ctcatctttaa gacttgaata agtacacaca 2040
gatcatcctt aaagattaaa tggcgtacgtt gctagagcat ttggcgtacggt aggaaacttc 2100

```

-continued

agatgcggtg	gctgagatct	cctgctgtc	tcaaggatgga	gagcatggtg	ccttgcacatt	2160
ctgtcacatt	cttttgtgga	ttcttatggga	tcaggctggg	gttccctcc	ttacacctcgag	2220
gccccatccatc	catccatcca	tccagtcagt	cactgtttgc	ttagagcaga	ttttggagta	2280
tgagtgaact	agggagagcc	tctgtgacac	caagagttaa	acaatagtca	caacaaaaca	2340
caatgcgcct	ggtgataagg	gtgagcttg	tggaggtggg	gcagaggact	catggaggcc	2400
aggagttggg	agatgtatgt	gacagccaaa	tgaggagctg	caagaaaaac	attctagact	2460
gaagaaaagtt	gggcgagtga	ggaagggagc	catgccaact	cttgctgcat	gcattcaggc	2520
agtacccaca	ctgctctgca	gggctgtttt	ccacagtgtc	tgccttcac	gagatacagt	2580
aaaagccctg	ggagtgtgga	ctgtacacct	acagggcagc	ttcttactgg	agggtttatg	2640
ttgtcacagt	gccaggggct	ggcttagcca	gagcaggaag	gcgcacctgct	ccaggcagcc	2700
agggctggga	gtcagggaga	agcagtctcc	tgagtgttcc	ctgtttctct	aaactacttt	2760
acatctgagg	tacagaaaact	gggttagaag	gaaaatgact	ccaaatcccc	agggtgtgca	2820
ctgcagactg	accttctgag	tccagggaaat	gccctggact	ctagctggct	ctctcgtag	2880
gttcacctcc	ctgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgcgc	taggggtggc	2940
acgggtgtggg	taaaactgtca	accaggaagg	accctgggt	aactgcata	ctcctgtctt	3000
aagaacccct	gctctaggac	ctaagttaga	gaggtagcag	aacctctagc	ctctgccagt	3060
gaaaagacag	aaagcacttc	tgcagggaaag	gtggtgagct	tgggtcaact	tgaactatgc	3120
agagaggcct	ggatacaggc	atagggttgg	tgggagcata	atgtaaatgg	ttcttagtgt	3180
tgggtgtat	ccctggggtg	caactgggg	ggtatgtgt	gtggctacag	tctgacatga	3240
ggtcccttct	agtcaagggt	tctgccccccc	gtgggggggt	gggggggtca	tccgatgtgg	3300
tttttaggac	ccagaaatac	aactttagct	gtagagcagc	caggttaact	gggtctgact	3360
ttgaccctgt	ggcctgaggt	ccctatgacc	actaaccgtt	cccttgc	ccctgggcca	3420
tgttgtatct	cagagtactg	gagctcaaag	atctggcttg	gagccctgat	tctacctttg	3480
ctgtgtggtc	atggtcaagg	gctggaccc	tccgcatgtc	tcttgggtttt	gtttgtttt	3540
gtttgtttt	tttcatttgt	ataatggaga	tgacggatgg	ggtgtgcctg	gacactctcc	3600
aaggattttc	tcaagtaatca	gggttagatgt	aaatagtgaa	gaaagcatct	gaaagtgaga	3660
tccttgcattt	gtgaggaaact	gtttaggattt	caagccaatg	agcagacagg	atgtctgttt	3720
gttcattttc	ggaggaactg	ggaggagtcg	ggagggtggg	catcctcagc	ccttgagggt	3780
taataaaccct	tcccgcttca	ggcaggatag	acaggcaaga	tcacattccc	ccctgcacag	3840
gctggcagtg	gggagactag	gccccctgect	gggggttcca	gctgggagtc	acaggtcccc	3900
gagaaaagcc	cttactgact	ggcgattgcc	attatataga	tgggcaagtg	gagactccag	3960
gattctgtca	tacagcccat	attccaaacct	tgctgttca	tcttggtcca	tatccttctg	4020
ccatgtggcc	ggcgagactg	tgctttgtat	cgctactggc	gtgtctggtc	ttggatagag	4080
gttgggtgaga	ggagtccctgt	ggggcaggtg	ccaggccaag	attttaggc	agccatggtg	4140
agtaagctca	gttggggcag	tgtctgtca	gcatgcacaa	agccctggct	ccatccctag	4200
atccgaataa	gcctggaaatg	ctggcacacg	cccgctctgt	atcctgecat	ttgggaggca	4260
gaggaaggct	atcctcgcc	gtgttagtgag	ttcaaggcca	gcttgggcta	agactgagtc	4320
tcaaaaaaca	aaaacctaaa	aggctgattt	atatctatct	aagccctgt	gtgggggtgt	4380

-continued

gcctctgggg	cattttatac	atgtctgtaa	aatggggatt	gaattgtcat	ggaccttta	4440
attgtgagga	tgcagtgtcc	attgctcagt	aatatgtgcc	acccaatcac	catectttcc	4500
ctctcttcct	ccctcccttc	ctctgtatgt	gcccagagtc	ttgtgagtgg	catgccttag	4560
gcagttcctt	caaagtcctt	gggagatagg	agaggagggg	cccagctgt	caccatgtca	4620
gcgatgtccc	ctgggttctg	ccctgagttg	tagctctgg	ggaggggggg	tgcagcagg	4680
cttggagggtt	gggggagcag	ctccagcagg	atggaaaatg	tggcgccag	ctgctgtgg	4740
cagaggatac	cacagagctc	ggcgcactct	gcaaggggt	gggcctgtgt	ccctggacac	4800
agccctgtac	acataaacac	atgcacactt	cttcagagct	caagcctaga	gaccacagt	4860
tactcatgcc	tggctgtatgt	atacattcag	tgactcaatt	atgctatatg	cttccagaaa	4920
ccagcctcca	gacacgttca	ctgtggatcc	acatggaaag	tgacagacag	acccttcatg	4980
ccacacacta	ccaaacacaa	cacttagcag	caggagacgc	caggctggg	agtacacact	5040
ccctcacaca	gcctgaagac	actctcagac	acaggaagtc	catatacac	ggaaagtgt	5100
cacatacgt	cacacagaca	ttatctatac	gcaccagcat	ccataggatc	ccagagactt	5160
gtcaggggac	acagacacca	tccactccag	tacccagtgc	caccttcc	tcagagagca	5220
agcaacctt	gacacaagtac	ctggggttca	acagatgtca	catgaatacc	cacaaagagg	5280
ttcgcaggca	ggatggaccc	aacctccgca	cacattagca	ccaacacaag	ggttgatcat	5340
gcagatgtca	gacatggtgt	ttgtttaacc	agggcaccca	gacatacatg	ctcatgcaca	5400
cgcacgctca	tttacactga	cggtgacatt	tcacctacgc	acataaacac	aggaactt	5460
ggtcagtgtt	ccccggagct	actgtgtatg	actcttctgt	ttccagctgt	ttcataaggc	5520
ttcttgtctt	gctgccggg	ccataagtgg	cgtcaggact	gggtttgtaa	ctgtctggat	5580
ctggggtccc	aggggcccag	ctgaggaggg	ctgcagggtgg	agtctggagg	tctccataac	5640
tggagaaact	gcatctaagt	tgcccaggct	ccagagccac	gccctgtggg	actaaggaa	5700
atagacacat	gtcaaggctg	ccacagaagg	aacggtgcc	caggtaact	cttctaattcc	5760
ctttggaaatt	tcctgcatac	cctgcctccc	tctgtgcaa	gtcacagagg	ccgcaatgcc	5820
ctgatctgt	gagggtgccc	atgcctagtc	tagaactagt	ctagaagtga	gaagaggtat	5880
acgcaccta	gagcagctga	gcaggctcc	tggaggagga	ggtgtcacag	ctggggattt	5940
gcctgtattgg	ctgcctcagt	gctcagcttc	tggataaggg	cctgcgcct	tgcaggggct	6000
cactaagtag	atgtgtatgc	ttaattgtatc	agctatggtt	ctggcctaga	aatggatagt	6060
aggtttgaa	aggaggcgtg	gagccagggt	tggaggtatg	tacgctcagt	ccttgtactc	6120
aggaggctgg	ggcaggagtt	caaggccagc	tggttactac	atagtgagac	ctgtcaggaa	6180
aggacagagc	gagaggagg	aaagaatagt	tgagtcatgg	gtgtgagagt	catggaacta	6240
gggggttgggg	ttgaggagta	gtagaaggtc	aggaggactt	aggaaggaca	ctaggtaccc	6300
gagcagtaga	cccttggctg	gaagatagtt	ttgttggttt	gtagcagggc	tggggccagg	6360
tggcccttagc	actaggcaca	aaattgctgc	tatctcagg	tgccaatgaa	ggcaaaagca	6420
aggaagactg	tgaggtecca	caggcacact	ccgtctttag	aatcaggtcc	gaggccaagc	6480
cacacccctgt	tatgtccctt	cctgtgcatt	ttgccttgc	gaacctcagt	cttccagtc	6540
gtatagaggg	aattcaatcc	cttcttgca	gatactctgg	gagataagag	ttcagacaca	6600
tgagagccca	tttctgtccct	gtatacagaa	gatggggaggg	actccaggct	ctaggtagcc	6660

-continued

tctgtggcag	aggcacactc	cagaaaaggg	cttggtaag	ttcaaggcaa	agggctggct	6720
gtcagccagg	gttctaactg	tggccagccc	ctagcctcca	acatatgaca	tcatctgttag	6780
tctatacaca	ccccctcggt	gcaggaaggt	ggggccccctg	cggtcagcac	tgtggctgcc	6840
tttgcgtggct	ttccaagctg	gggcaggagt	cggaggaggt	agaagctact	gttcatccct	6900
cccttgcctc	agtctaaagg	tctgaaagct	tgggatctgg	accagectg	ggctgcatact	6960
gggcctgtgc	tttgctccct	gttccccctc	cagggaaatgt	tctgcagtgt	caccatgatc	7020
attacaacagg	agggagccgg	tgtgtgagga	gttgggcccc	tctgtgttcc	actcatctg	7080
gttgcagcc	cttacccttc	aacagctcca	gaaatgtgt	gacacatttc	tccttggcag	7140
ctcaattacc	ctaagtcaact	cctccctga	tggccaaaac	ctgtgtctca	tttgcacacct	7200
ggtgatgctc	gcttagtggt	gaaggaccac	ccttgggggc	ccattccaag	atagtggtaa	7260
aacatctct	gcctgggcta	cagatgaaaa	taaaggtgac	tggtgtgtgt	gtgtgtgtgt	7320
gtgtgtgtgt	gtgtgtgtgt	atgtgtgtgt	gtgtgtatgt	gtgtgtgtct	tgattcccc	7380
actgtggtg	tgtatgtgt	tgtgtgtgt	tgtgtgtgt	tgtgtgtgt	tgtgtgtgt	7440
tgtgtgtgt	tgtgtttgt	ctgtgattcc	ccaaactgcct	gagcctccgt	ggctttttgg	7500
gactcttcag	tagcaccaga	ctttgggg	tcacccttc	ctggcttgcc	ttgccttaggc	7560
accatccctc	tgtcccagcc	catggctaag	tgcgggctg	agtcgcctt	aagtcttgct	7620
tccccagcct	tttcacccag	acgeccccag	tgtccctttt	gcctttcat	tatggatcca	7680
ttccccctgtc	ccctttccat	gatccctgg	ggactgctcc	aggcctgggt	tcctgtgtta	7740
ggtaactagga	gctgtgggga	ctccctcggt	ggacatttag	ctcacagtga	ctagacttgt	7800
taacaggaca	tttctggctt	cgcagcagcc	actcctgctt	ctgcatgcag	tgtggcgctt	7860
ccaggggtgct	tagccccctc	cttctaaacca	ggcttaaact	gcagcccgag	cattacgtca	7920
ccttgcttgc	tagtcacatt	agctgtaaag	gacaagaggg	attcaggccg	tgtggttctc	7980
ggtgattctg	caagcctcga	ctccctgecc	ttttctgtta	cagacttggc	tgaagcccc	8040
aggttgggg	atgttgc当地	caaaccctga	tccaaagctg	ccttccttcc	tggcaggagac	8100
aggagagaca	gtgtctgtgc	ttccttgagt	gtgccagca	gcccagggcc	tcggggcgct	8160
gccccttaggc	tacacacaca	cacacacaca	cacacacaca	cacacacaga	gagagagaga	8220
gagagagaga	gagagagaga	gagagagacc	tggagacaac	ctcaactccc	tccttcagcc	8280
tcagttggtt	ccccctggaa	gtggccagg	taagaataga	gcttgccaca	aagtatcacc	8340
aatgagaaag	agagaatagc	ttctcacctc	tgctgggtgc	tggcacatgc	catacccttg	8400
gaagtgc当地	gctgttggta	cttctgtgt	cccatttatg	agcctcaaaa	tgagcctaag	8460
tcctgactcc	tgacgcctct	gctttcaggc	ttcttgccata	ctggggagg	agggactaac	8520
tgtgtgtgt	tggcatgagc	ctggccgtgt	cagtcgtct	gctctctct	gtgtggaga	8580
ggcagggccc	agagaagtgg	gacctcatag	atctctaggg	tcaagatcc	ccagtgccaa	8640
tggAACAGGA	gcattgagat	cctgttaactc	ggtgatgcat	ggctggatgt	gacttcctct	8700
gtgacactta	accctgattc	ccccaaacct	ctgcccacag	aggctaagct	gggttgagct	8760
tgaacaggag	gcaggaaccc	agagtccccc	ttggcctcac	tgcgcgtct	gtaaaggaca	8820
aggagtctt	gagttccct	tctccctggag	tgtttgcatt	caaactggga	gggaagtgaa	8880
gcttttgca	gccttctccct	ggggctctgac	tgtgtatgatc	tcatctggag	ccctgtcaga	8940

-continued

gtcccgaaaa ggaggaggga cagcagatgg ctttctggc cttttgggtg gggctggct	9000
gctgtccctg tggccactgc tggtcagetc ttccatagct ggggttagtt gccaggccc	9060
gagtggggtga caagtggctc ggagccaggg tggcagatgg gtggcagcc tgggccttgg	9120
cccagggggag gaccatggta gccaagtggt gtgggaggag gagggcaaaag cggagcccc	9180
ggtccttgt agagtggtt attagagtga agtccctggc cccatataga gcatagcttt	9240
tcctgtccat ccctaagtct gctttgagc tggcaggaa atgatccatt tggtaagat	9300
ttggaatatg tcacggctgc ggtcgggtgt catggctgc ggagttgaag cctccgtgg	9360
gattggcctc ggagtctaaa cagttgctgc agctgctcag tggatgggtg gtcctccett	9420
cagcctggct gttctgttcc ctcatggcta tttcttgagg ccctctataat gtcccgttc	9480
tctggtgact gaagtggttc ttccacaacc aagcagacaa atagataaaa taccctgtcac	9540
taatagaact gtgtcgtgca gtgtgaggga aggagaagg taggctgcac ggcttggcag	9600
gaggcctgac aggaggtgcc ataggccacc cacacaggaa tggcagaggc tgaaggctgt	9660
gacaggccag gtgatttctg ctggcagg aagactcggg tggctgagtg tggecccaeg	9720
gaggggtggta taagggacat ggaggtaag ggagtcaaggg ttggcctgg tggagcctga	9780
gctccttagga taggtttgtc ctgcgtatgc aacgagaccc cccggaggtt ctgtttctgc	9840
tttgctttat gtctatttga aaaaaaggc tcactgtaaa ataagtttg ataaatgtac	9900
tataactgtt aacccctggc aagataaggg ctaacgccta taaaactaggt acattaggat	9960
tgtatttctt cttcaatcta tttatctatc tacttagtta attaattaat gtgagtgtgg	10020
gcgtgcgtgtt gggggctt gtggacttg gttctccctt cccatcaagt gggcttgg	10080
aattaaagtc aaatcgtaa gcttgggtgtt aagggtcttt gctccctgaa ccagctcaca	10140
ggcccctaat tggatacatt aattttttt aaatttaatt ttattttac tgtatataag	10200
ttataacctga gagctaggga tggtagtagct cagggccaga gcctggactt cacatgttag	10260
agatgctggg ttaatccctt agcaatcaa aaagaaaggg agtgggggat tagaagtgtt	10320
cattgttcag ccaaacttcc tctacaacca gagccctccc tcgactactg atctgttctc	10380
tgtcccttag cctcacccat tcacagtgtc acatgcattt caccctcagc atctcctgtg	10440
gctggcttcc ttgcaggagt caaagcccg ctgctccagg ctcccagac accccacccc	10500
caccccccacc cctcgctgtc gctgggttta catgggggg tcacagtttta gccattttgt	10560
agtcaaggagt cacgtggct gccccacgt tttgatttca attggaggc taaaaggcagg	10620
gaacatttaa acagtttaa tcacatttttta agtattaaaa aaaaaatcac aggctggaga	10680
gatggcttag gggtaagag cactgactgc tcttcagag gtctgttttta caattccag	10740
cgaccacatg gtggctcaca accatctgtt acggaatctg acggccctt ctggccgtt	10800
ctcacatataca taaaataaat aaatagttca taaaaaaaaa atcaactttt gggagctggg	10860
gtacagagca agcactaata tgcaagggtgc tgagttcaat cccggagct ggagaagcga	10920
tgggggtgggg gggagaaaga gagggaggga gaattactgt gtctgttact aagttgggg	10980
cagcttactg aggtcaagga agggtcacca tgacaagaat ctggcccg ccaccaggcc	11040
tgctgttaggc actcaactaag tattcaaaca cctggatgtg cccaaaggaa tctgttttc	11100
tgaggtaaag agttaggagg cagtgacaga gcaggaggcc caggtgtgct gtctgtggc	11160
tcactgcgtt aggtcaaggc ttccctgtt ctgaccctgc ctggccacaa ctggatctc	11220

-continued

-continued

tctactccct	gtctgcaacc	tggacttaga	tgttgtctaa	ggcacagagg	gcagcaggg	13560
gtcagagcct	ggcgggcgt	cttgcactct	ttcttcagct	tccttaggt	tgataacagt	13620
ctaggggtct	aatttatttc	agggtgttag	gttgagaact	ggggggccct	cagactccag	13680
ggtgggca	ctgcacagtgt	gcctttaca	tctgaggtag	agggtgagat	ttggagtgc	13740
cccatgctgg	gaagggggaa	cctcagaacc	ttgttctatg	agattgtatcc	ctcttcttg	13800
atgtcctata	gaagagttca	agaggtggca	tctctgtggg	gcctctggac	tgtctgtgaa	13860
gaaaccagga	ggcgtgtgc	tcagctctct	ccaactggac	ctgtcacc	agagcctcca	13920
aaggcaaaac	tctgactgt	gggcccaggt	gcccagaaga	tctctggct	gggtgtggac	13980
accgagcctc	tgtctttcc	ctgccccctc	ctcagaccca	gataccaga	gtctctgcct	14040
aagcactaca	gtgtccagcc	catgtgggt	acaagcctgc	ttttgtcaac	tcaagccgt	14100
ccaggagggg	ctgggtgggt	tttgcccg	agctggggca	agccccaga	ggcttcagcg	14160
cccccaaaaa	ctcagataca	gaactggcat	gcccacaaca	gcctgcact	ttggtgagg	14220
agtcccccc	agagacagct	gttggggata	gccattcagt	cacttctcc	ctagacgcct	14280
gtgtcaggct	cagagccagg	gacttgtg	gtgagatgg	gtctttgtct	tctgtgggg	14340
agaaaagtgg	aaggaagt	tttgcctcg	aatcaactgg	ggaacccacc	cctgaagtt	14400
ctgcttagta	ggtgtggatt	ggaaatatgt	gtgggggtgg	ggttgcata	ccctgggt	14460
gaaccacacc	cctgagcat	tttggcagac	actgcctgt	agctatacca	tcctgggg	14520
cacttgtcat	tcagaaccag	atagagaaat	tgtgattgt	gtttctgt	ttttcaaaa	14580
ctttctgttt	gaccgtgg	gactaccc	acaaaagccc	gttacttaaa	atatagggg	14640
atgtcacctc	ctgtgatct	tatggagc	catctccacc	tgtgatccc	tcagcctgag	14700
ctttagtat	tctgtgttcc	ccttgactca	ttctgaccta	gagtcact	gaccctcaca	14760
tcctccca	taccgtgt	tcaggataac	aggtaacat	gtgagctgt	cgtgggtgg	14820
aaggccctgc	agoctggcc	atggatgt	ctgtctgc	tcacttct	cccttcctt	14880
cctgccccct	tccctgccc	ctgagctgt	tcatggc	ctcacttccc	catttctccc	14940
accgactgtc	tcacttcc	cctgtccac	ctggctgc	cagttggc	gtgtgggt	15000
ggtgactctg	ggtacttga	gatgactgt	cctctgc	cagtaggc	ggtcacccac	15060
tccgggtgt	ggtaacaca	ccctctgg	acagtctgt	gtgcacagg	gatacttgg	15120
tcagttgt	gttgagagag	acagcatgt	gtgggggtgg	gccgcca	tgtcactt	15180
gtttccctt	cctgggttct	gagaatagt	tccagaagtt	aggcagatgg	atgggggg	15240
ccctggagt	tgtgtgg	tggatagtt	ggcttcttac	ctgtccat	atccagca	15300
ccacaccc	tcccctcaca	gggagccatt	gcagccag	catattctgg	tcccactg	15360
gtgtggaccc	acagcctgt	ctcgcccc	tggaggctgg	caggcagg	tgaggagagg	15420
tctccgtgt	tcttttcca	gactatcaga	tccctgggg	gagcacagg	tgtgggg	15480
agaggggatc	ccacccctac	accctcc	ttattccat	cttctttag	ggactcc	15540
cctccagagt	agocttctg	ttctgggt	ctccctgt	ccatcag	caaccttgc	15600
agcttctct	gggagcagag	cctcgt	ggcctgg	ctgggtact	agttcc	15660
ctgtgtgt	aaatcgtaaa	gctaacagca	aatattattc	atgtttgt	ctcagttc	15720
acatttctta	tgagacataa	ttatgtgtt	cctctggag	gcttccacat	ccccctctg	15780

-continued

tgttagtgtct gggagtcacg tcccatctt cacagtgaga agtctctttt attctctttt 15840
ccttagctag gcatgttttgc tttagtacgtt tttagagatgt gcaggaggct 15900
aagcaggtgc tgggtgtgtg tggggggggg gttagggcag gcaggttaggg cgacactgaag 15960
attcccacag tcactgggag agcaactgcc aaacagcttca agatgagctg aggtgggtga 16020
ggcacccgata ggagctgtga cagaggtggc agtagtggg cttagctgagc gaggtcacct 16080
gaagatcccc cactcaaaac tccaaaccaga cttccaggg ctgcaaggct gateccctcca 16140
tatccccagc cttgtgggtt ttggacagtc tcacatagat gaggctgggt ttgaaatcat 16200
gatcttctta attccatatac cccgtgcctg acccttgctc gatctttaaa acaaacaac 16260
aaaagctttt tttagacaata gtcccaactat gtataagact tggctggct gcagctttct 16320
ctatagacca gactggcctg gtctgaaatc aagtcaactt gctctgagat cctgcctttt 16380
tgtctctgccc ttttttcaga tgctgagatt aaaggatgtc accattatta aataaatatg 16440
tttttgatta aaaaaatttt ttttcaagac agggttctct ctgggttagtc ctgtctgtcc 16500
tggaactcac tctgttagacc aggtggcct tgaactcaga gatccttctg cctctgtctg 16560
tctcttgagt gctggatca aaggcgtca gcacactacc tgaccaatat actttttaaa 16620
tgacattttct ttgtacatgt atgtgccaca gaatacatga ggagatcaga ggacagctta 16680
cagaagtggg ttctctctc tcttctacca cgtgggtccc aaggaccaaa ctcagccac 16740
taggcttggt gggcagcacc ctgaccatc ttacctgtcc cattgtatgtc ttcttttttt 16800
ctcagctgag acaggccctt actgtgttagc cctggctatc tcagaactta ctctgttagac 16860
caggtggccc ttgacatctg cttgcctctg cctcccaagt gctgggttca aaggcatgca 16920
ccactacact caacgtcttg cagtgctctgg atacatttagt ttctacttct aagcctttgt 16980
acacactgtt ccctcttctc ggaactccctc tggatgttgc ttcttgccct ggcttagatg 17040
tcccttcttt agagatggcc cccttcctcg ctcactgcca tgtaaccac attgttactt 17100
tgttatTTTT gcctccatga tacttactta acagtatctg aatttgcattt ttgtttactt 17160
tgTTTCCAA acaaggctgc gctaacaatcc ttccctgggtt ttgtgtgcta cacagtgtcg 17220
gttatgttagc agacactcat gaatgtttgt ggagtggatc cctgggtat gttccctt 17280
gtttctgtct gcgacaagcc tctactgttc cctctgtcgat tatgacaaga gaaggaaggc 17340
tgtgtttccc atgctttgg aggctccgt tcaagatcat tttgttctc ctgtgacaca 17400
gagccatcaa ggcaggggaga ggcgtgacgca ggaggctgtc acctcatggt gaccaggaag 17460
caaggcacag ggtatcaata gccttcagg ggcctttccaa aagatgtct ctcccttgc 17520
tcctgaaaat ggtccactgc ttccacagtag aaccatgagg aataactcagc ccatagcaag 17580
tgactttggg gggggggggc gcttaagat ccaaatcgta agagtgggtt ggctgtaaat 17640
tatcttccag gtctttgaga ctcgagtcctt ggtggctcac atgtctgtc agtgttctga 17700
gacaaacagc atcctctctt gggcccttttgc gccccttgc atgggttcaata ctgtggaggc 17760
cacctgtact ccctggctgg tgcccttccat tctacactgca caacactgtca gtatttcccc 17820
tatttctccc tcacattgtcg cgctccact aacttccctc tcatggccca cccagatggt 17880
caagaataat actcccatct agagaccctt ctcttagtct cgtttgcata gcccccttt 17940
ctgtgaaagg catatattta ggatccctg aatttagctg tggctgtctt tgggtgggtgg 18000
tattcagcct gtaacccag ctggctttctt gagtggtggcc agcattcagg caccctggct 18060

-continued

actgccatga catctgtagg tctgaccctt tctttcacc cctcaggatg agattgagga 18120
gctacgggct gagatgctgg agatgcggga tgtctacatg gaggaagacg tgtatcagct 18180
gcaggagttg cggcagcagc tggaccaagc cagcaaaacc tgtcgcaccc tgcaatcc 18240
actgcgttaag gctgagcgcc gcagcctcg cgctgcccag acaggccagg ttgtatggga 18300
actcatccga ggtcttggaa aggacgtcaa ggtcgtctg gggtcactgc caacccatcac 18360
aggtgccctc gttgggttgt tggatcataa accccattgg ggatccaggt taggcctgtg 18420
gcttactgct tgccttgagc aaattatcct gcctctctgg gcctcagttt ttccatcttc 18480
aaagtggggc cccctgggtgg atcaacacat cctccaaaca ggatagccag gaaatggtat 18540
gacaagaacc ctctcgaagc agcaagaggc cgtttcttgt atgtctgtt tgtagtggc 18600
tttggctct gatggggccc tacttatggt ccctgctcag caccacaggc ttcttatggg 18660
aagactgcag ctgaaccagg atggttcaga gctgggtcgg aggatacttg ggctttgagg 18720
tttaaatcaa gtcttctata tttcaatgtc ctcatctgc aaatgggtat aagaacctac 18780
aatccagaat attaaaaccc tggctaagac accagaggga ctgacagtga gtagctatta 18840
ttactgcac cgttttctgg aggagggggc tggcctcag aggagagaag caccatcca 18900
gggtcaagac gtcaggccca tccacccctc agacactact gctggaaacct gctgtttctc 18960
tgcctaagga gaaagcctga atggcgccac ttaggtcggg gatggaggag agaagtcccc 19020
attgccagtc tgtcttatgt aageccccaa aatgtggggc ccctgagccca gcctccgctc 19080
tcagaactca gtccagtgtg aacaaggaac tcaggggtgt gcgaccggta cagaactggg 19140
aggagcctag gagttctgca ggttagccaa gggggaaagag ggctgtgagcc tgggaaagat 19200
ccaggacttc cttggggaaa gatcccaagg gccccggagg gcaggcaggt gctgagcaag 19260
ccacacaccc gctgctggaa agccctgaga atccagccct gcagggcatt cagaaaggac 19320
actgttcatt ctgtgaccag cggttagact ctatgcaccc ctgtgaaacct gtgtggagcc 19380
agtttccttg actgcacact aagatcttgt tatccaaaaa atcataaaca aaacacgatt 19440
taaaaaaaaaa acaacctaac aactccccgg ttccttcatac ggttaggagac agagcaggga 19500
gggaaagagc ggagcgagcc gttcaggcga atctgcgggtt gggccctggc ctttgaatct 19560
gggagggacc ctagacaagc aagaccgaga agagagggtgg ggagcaggct ggtggggcggg 19620
gccagagcag ctggggatggg ggtggggctc agccctccag gaggaggggg agggccccc 19680
ctgttaggat ttgaacgagc caatctgcac gcccgcgtc ctctgcacca ccccccgcac 19740
tgcaacgcgtc agctcatgtt ctggctctgt cttagcgggc cgcacagttt tgccatcacc 19800
cttcctccca gtgtgtggg gatgttagag gcaattatgg aggtcctcag aatgtctact 19860
tccctgtctt ccccttaggca cctctcaacc aggtggagag ggactggggg cggggacgg 19920
gggggggggt atgccttaggg atccctctcc tgctcagcag gctccttggc cctgtgcctt 19980
gcagggcttt tgagctccgg tggttttatt gagctggat gttatttagt agattccca 20040
gccggccaccc cagaacaacc ctcccaggct cttttgtgt gtcacacagc ttaacaagat 20100
ctccgaccct caacaggag acccgctagc atgaggatgg tctggccact ttatggccgt 20160
gtgggttgg gtaagctgtc tcaatctctg atcctgttcc cactgtggaa catggagata 20220
agtttagtat cctatagaga cttaaaaaaa aaaacacaca aacaaacctt tatctactat 20280
gtctgttagct ttctgcttac atatatgcct caacaccaga agaggcacca gatccatgt 20340

-continued

-continued

ttttgtgact cagctggcag ggtgcttacc tagcaggcac gaagccctgg ggttggctt 22680
gagtgcgtca tagactaagc aagacggcat agactaaagca agatggcata ggatttagcac 22740
agcgcttggg agacagagac aggaagatca gaagttcaag gctatccctg actatgtca 22800
agctcaagtcc ctcctggag aggaaaactt ttttgttcc ctggagatgt agttgagtgg 22860
gcgagttactt gccttagtgc cacaaaagtc taaccattcc cagcacaggc taaaagtcaa 22920
aatccctgtg tggatggctc cttgacctcg tgccttgag ctgcccagtg agcccttaa 22980
tatTTTACAC aatcaaaaac aggtcaagaa actatgaaat caatacacaact ttgttaatac 23040
actaaataga ctgaattatt taaaataat agtaaataag aagccttgc tttttgagc 23100
acctcctggat tccagccac cactcaagca cccatgatgg catttgcctc ctctggcac 23160
cacgggaaaca gtgtatccat gtaatggta ccagatgtag gctccagatg ttccaggc 23220
tggcactggg tttgaatcga gttttttctt ctgcaaaatc ttttcttgc tattgaagg 23280
tggggagagc tttgtggccc acagtggcca cagggggagc ttgtggccct ctgcctggcg 23340
tcctcagaga gctccaggct agagggggca gtggggccag gagagagtag agctgtgecc 23400
agtttccctt caactattat tttccctggc cttccctgc cattgccaaa ggggtctgtg 23460
ctaagttggg cttggggcca ggcagaggc agagccaggc agggccttcc cccttccaga 23520
gggaggagac acgtggtcat ttccgacatt ggcagccttg agtgcttagt tgtctgtgcc 23580
acacaggctt ggctgaccag gctttgggc tcctcagggg gctctgacta tgccaggctcc 23640
tccaaaccccc agggagtcct agcccagecc aggcattgccc gatgtacagg accctgagaa 23700
agtacttctt gctctgacact tcagtaaagc cttagcatcta gtgggttaga gtttagcagg 23760
cccagcctac gtggggagtg gcaccttattt gcatatatct gctctgttagg gcttccttca 23820
ccccacctca gcccccttgc cctgctttcc cacgcccagc agcagggcca ggactgaggg 23880
gatcttccca attttgggt gctttccctg ctgagcctgg acctgaaatt ggagactgct 23940
gttctgtgag ctgcataa gggatggctt ggacactago ccagcctgtg tcattctagc 24000
atgttcttta tcccagaaca ggcctcaggg tcactgtgac ctccagcaaa tcatgcccgt 24060
ctctggacct cggttccggg gaggcagcgc cacgagtcac cagacactca ctgtggcagg 24120
cactgggtga aggctttgtt agcaatagcc gatgtatcc tgctgtcggc cctgagatgt 24180
ttctgtact gctgtggcca ttttatacgt caggaacact gaggcacaag gccaggaaat 24240
gacttgcga gaagcacaca ctgccccgtt tgtgctggcc ttccaaacca gttcttttc 24300
ctttcaggct cttgtgaaga cagggctggg agtactgtt agtggaaagag tgattgtca 24360
agagaagcaa ggctgggttc agactccagc acagctctgg cggAACATAA aggtacttt 24420
aagaactaaa ccagaggggt tggtgagatg gtcatttgtt taagagccac tgctattcca 24480
gagggtcctga gttcaatccc cagcaaccgc atggggctc acaaccatct ctggaaagt 24540
ctgatgccct ctccaggcag gccaatgtat atgcagctga gcaactatac aataaataaa 24600
ataaaaaaaga aaaaaataaa agatcatcac agtttgggg gggcaacata ttgggagctg 24660
gagtgcttgg gatcgatct gataatcctt agattcatgg ctcttggta gacttttagac 24720
ctctctgtg agtgagggtg aagaggtgac actgttaacca aggagagatg ctcactgaga 24780
ctcgagggtgt gtcctctcc tgccttgacca tgcctctgt ctctctccca ccctggctca 24840
ggtctctaag gacatctcca tgcggcttca caaggagctg gaggtgggtgg agaagaagcg 24900

-continued

```

gatgaggctg gaggaggaga acgaggggc tcgacagagg ctcattgaga cagagctggc 24960
caagcaggtg ctacagacgg agctggatcg tcccagagag gtgagggcgg cagcccgat 25020
ggggctggtg tgatgcatac ccctgcctgg ttgcttgg actgaagacg ggcacccttg 25080
tgccttgaca gcgctgttcc tccatccacc cttccagcat ttgctgagca cctaattctgt 25140
attcagccct gggtactctc tctgggacta actgaggcta tggcagcaaa tgcccgat 25200
gggcctggcc ctgtgcctca gtttcctgt ctataagact tggcagttgt tccctggta 25260
tcgacagggc cactgactgg cttgtgtgaa gatgttttag ggactagtgt gctgtggcca 25320
cttggtttaa gaaggatgct aggggccgggc agtagtggct catgcctta atcccgac 25380
ttggaggca gaggcaggta ggtttctgag ttcgaggaca gcctgctaca cagagaaacc 25440
aaggctatac agagaaaccc catctgaaa aacccaaaaa aaaaaaaaaa aaaaggatgc 25500
taggcttggg tctcccaactt ttgtctctga atgtgacttt cagatccatg tccccagtc 25560
tgggtatgtc agcccttcct cctactgtaa ctttcctccag ctgtgttta ggctcacata 25620
aagacagttt atctcagaaa cctccgcaaa ggcccaacaa agctgtggac attccatggc 25680
cgagtgagca gcagccggcc aggaagagtg aggccagattc tcaaggatcat gcagggtgca 25740
ggggacttca cagtggaaag ccatcagccc gggtgatgtc taaagcctgg gtaactcctg 25800
attcctggtg gggtcagcag ttcaaacaact aggtggcctc tggccccaga ggcatacgctg 25860
gcctcagct ggttgggtgc tcatgaggcg aaccgggctc cattttctgg gggcaccac 25920
ctgcttccca agaggcccg catgccagac gttcaggact gagacatcag tggccccgg 25980
gcagagcagt cttttggctg ttcttccatc tggcatgtt ggtgctatgg gccagtgcca 26040
tgttaggggtgg agagggggagg acctgcccctg tggccttctg tggagggcac catgctgtgg 26100
agctgggtgg gaggcttccct gctggcttcg ggctgcctat gcactgcgaa ctgtcgac 26160
caaccttgc tctttcatct ctgatggtaa aaaggagggg cattacaag cctgttcttg 26220
cacagctagg gagactgagg cactcggtcg cagttgttgc agtttgc 26280
cagcagagac atgtctgcag ccacttcccc ctccaaagtcc cggcttcttc ctctgtgatg 26340
tcagccaata gtacgacgtt aagaggatgt ttggatcac taagaagttt gtggAACGCA 26400
gcggcagggt agagcatgtg cttggccatcat gggaggccct gggttcaagc ctttagggcc 26460
gaaagccaaag gatcaatatt gtaaaatgtcc ctttatctga gattccccag agtagtgaat 26520
gagattcata agacagggag cagactact gggcgggggaa ggagggagcg aggctcactg 26580
gggggtggaa tagagttca gtttgggtgt atgaaagggt ttggaggcaactgtgcctg 26640
tgatctaagc aggggtgtttt gttgtacca acacctggaa ggccaaagca aaagacttag 26700
gaattcaggc caccctgggg tctgtgagac cctgtctta aaaaaaaaaa ggcctggttc 26760
aaggccggca gtatgggtgc atgcctttaa tcccactact caggaggcag aggccaggca 26820
atttctgagt tcgagaccag cctggcttac aaagttagt ccaggacacg cagactaca 26880
cagagaaaca ctgtctcgaa aaaaaacaac aaaaaacaaa atagaatgtt taaaatgata 26940
catttatatc atctaaccctt taccacatgt taaaaatgttgc attgtatgt 27000
caagctcccc aagggtttct gtaaaccaga gaaggaaatt aagaaacttag gcctgggtgt 27060
tgaagggggaa gggttcttaggc ctgtgtgac tgcagttgtgt ctgtctgtct gtctgtc 27120
ctgcctctct gtatgtgtgt atatttgggtt gtgtatgtgt atgtatgtgt gtatgggggg 27180

```

-continued

ggtatatgtg tgtgcctgct tgtctgactg tgcaccacat gtgtcccggt gcctgcagag 27240
gccagaagat ggcattggaa tccctggAAC tttagctatac gtgtttgtaa gctgctcagt 27300
gagggagactg ggacgagagc tagtcctctg caagagcago aagtgttcc acctgccaag 27360
ccatctcttc aggtccctgg cggttaaata agaagaatgt acagtcccag tagttagaa 27420
ccaactgggc tctgtatgtt gaggccagcc tgggtcacaa cgtaagactt tctcaaaaaca 27480
caaagaataa aataaaaatg gattggcact gtagctcaga tggtagagcg cttgcctagg 27540
atgagtggc tccctggactc aatatgcagt tcccaataaa atggcgctgg ctccacacat 27600
ctctacccca gcacccagga gatagagaca tgacctacag tagaagctt gggcatacct 27660
tttgtggaa atccctggct acatgaggca gtgcggcaag agaaagtcc taaacaggac 27720
atgatgagaa cgtctacagt cccgttacag ggaagcctga gggagaagat ggctgatgtc 27780
taggtatcta aaagctgctg ggatagcata gcaagaccct gtcttacaac aaacaaacaa 27840
acaacccat gatTTactta ggaattactt gtgagattaa aatTTacaa catttgaaaa 27900
ctttatTTca tgcgtgtgtg tgcgtgtgtg tgcgtgtgtc tgcgtgtgtc 27960
tgtctgcata tgcgtccac tgcgttgc tggcaatcag aggacaactt gaagcactt 28020
gttctttct tctacactt ggttccatgg gatccacac aggtcatcag ggttggtagc 28080
aggcccttt tgcctactga cttctcacca tgccttttgc tggaccttta aatgttagatt 28140
ttagggtttt aatgtaaaga tcccaatacg gaccctgaca ccgagtaggc gctgaagata 28200
ttgggtctca ctccctggaa gcctgtatgg gctcccttgc gtttagtgc ggctccagag 28260
cctgttttatt tgcgtttgc cctctcttgc tccctcttgc tcctgacatg gttgtatgg 28320
agacaagggt ccctcagcac ttctgccaga gtcacagtct ttccttggg atctgcctt 28380
tccatcctga cagatgtatg ggtgtatgtc tcacccact gggatggga gctgtatgtc 28440
ttgcgttggaa gacagtcctt ggcgagacag aaggggagggt gcatgttgc actgtgtt 28500
gcagtggcac cgtggccag ctgggttgc gttcagaat ctaacggcag ctgccttgc 28560
atccccacta tcatctccct tgcgtgtcagg cagagagca gtgtgagagt gtgttgc 28620
ctaacacagc cgctgtgtct gagagcttac cctgttccag gcaccactgc cagcccttgc 28680
catgtgccgt catggcggtt aaccccccgt tgcgtgttgc ttgttaggtac tactccttgc 28740
acatTTaaa atcacagttt ttaattcagt gatcacagaa tgcaccattt tacagttcag 28800
ttgcataatgtt tgggtgtctt ttatcagcat gaccgaattt cagaaagctc ctttactct 28860
tgtgagtaac cccacacctg tttagcgtca ctcctcactc cccgttagccc ctggcaactg 28920
ctggcgttgc ttctgtctt aatgatttgc ccaatcagga gactgcatta aatgggatc 28980
aagcaatgcg tgcgttgc tgggtggcgc tactacttag aacaaggctt tcagagtttgc 29040
gacatccatg ttgcgttgc tttttgtttt ttgcgttgc tccctatgc 29100
tgggtaatgc tgcattgtgtt gggcagatgt ctttataattt gatatttggg tgggttgc 29160
ttttccgttgc tccatcaataa tgcgtgtgttgc gacatgttca agttttgttgc attctgtatg 29220
tgtttttctt gggatgttgc acaggagtag aattgttgc tgcgttgc aatgggaaaga 29280
tacgttaagggtt attcaaaccc cggccctgg ctctccaca ctcacactca ctccaaacag 29340
agttggcgat cggctccgaa ctcatcaagg ggcttccag agcctgcattc tgcgttgc 29400
catctggcattt tcccttgc tgggtggcgc gaggcacttgc agccacttgc ggttggcct 29460

-continued

```

ggttctcctg ctgtgcagtg tggctggcc ccagtgtgtg agctgcagag caaaccaggc 29520
caaggagcca cgtagacctg ggcattggaaat cttggctttg ctgttcctt gagagtgact 29580
gccacccctgg gcagatggca cagcagctca atgctcaattt gctctgttta cagaaaggga 29640
aaatcagccc ctacagcttg tgggtgtttt tgaagattaa atgaggctgg gaagtgtca 29700
gcatagaata ggccgcttgct aaacacagtg tgcatttaccat ctcaccaggc 29760
caactggctgt gggcctggag aagecccttc tgcctccctc gcacaggggg tccccccaca 29820
gagctggcag ctgtacgtgg aaggaacagc tacgtgtaaa actgttgggt cttaaacttat 29880
tttttgcgtgg ctttgggggtg actgtaatgc tcagagacct tgctgactca gcatgatgat 29940
ctgagttcag atctccgcag cccacataaa aagctggca tgatccaaag cagacctata 30000
actcccgacac ttgcgttgcc tccaggctca gtggaaaacc ttgtctcaaa ggagcaagg 30060
gacaaagcag gacacccgac atcctggcatt ccatgtgtgt acaagtgtgt ggcatacac 30120
cacatctgca tacatacaca cactatacaa tttaaaattt aaaaataaaaa attaaaaat 30180
tttttctcc acccaagtgg tggctctcc aggggactat atggcttgcc tgcattggcat 30240
tttggccatg gtgcaagag catacctcac tcttagagatg ctcccccac ccaggccca 30300
ttcatggtca gttctgtca catagcaatc tggtttcaatg aagacttaca tggcacacc 30360
ccacatttgc tgagtcggca gccaggaagc tcaggagata ttggggacag aggggacaga 30420
gtttagacag ctctctcctg aacttagggc actgcctggaa ggcttctggg gaagacagac 30480
cccaggctca tgctgggaag ggttggccca gggattctca gagctatcac ccaggctgt 30540
aatagtggct tgaggagtttgg ggaatctggcc ttggggctgg gtctctggg accttagacc 30600
tggagaatcc ttttgggtt cttgcattt accactctag gattcctact tgagaggta 30660
acacacactt cccataggat tttctgtgtt ttcacacaca ttatttcta tgtatatgt 30720
tgtgtatgc ttccctaaat tttaaaatgc agtataaact taatgaggca aaaatctgc 30780
cacaaaagcc cataaacgcata tacctgacat gaccctgagg aaagttccag caaattccga 30840
ggcaggaatc cgggacactgc tccatcttc tggctctt gtgtactctc tgaatcttc 30900
ccacgtgact gaggaatctt attaaaaatg acagatctag cttaaacata tacaaaataa 30960
tgaaaggcctt cccaggccac cccctgtctc atctgcctaa cctgaggata acagtttgc 31020
ttgtgtttt ctggatcttt ctctgtgttt gacaatgtgc ttttctgtgt cttaaaaaaa 31080
taaagatggg ctgcgtttgc acacagggtt tctgtatgt gcaggatcat gttgcata 31140
tatagtccctt taaaatgcca gatgtttgg gggtttctt cagtcgtaaag tgcttattat 31200
agaaaacaaac aaacaaacaa aacaaaaacc acggaatcaa ataagcatag gaaggaaaga 31260
aagtcataaa tagcactctc tatttcaacc ttaaacaggg ggctgaagggtt tagagtca 31320
ctcctgaagc atcagttggc aaagatccct cggggagggg aggaggatgg gacagcatga 31380
ggggcgtggc tttagctacc tgacttatgg ctttgcgtgg tttttgtt gtttagttgg 31440
ttggttgggtt ggcttggggaa gggggatgtt tttatattt gttttcagac acagggtttc 31500
actgtgtatc cctgactgcc cttgaactca ctctgttagag caggctgacc tgcatactca 31560
agatctgcct gcctctgcct cttgactgttc agtaccacca cccagcaaca tatgtgtttt 31620
ctgatgaggc ttccctggaa ctaactgcac accaggttgg actggacttt ggcatacaca 31680
caagctggcc ttgaacttcc ggcattccctc cagtttctat ttccaggtgc tggaaattaca 31740

```

-continued

gatgtgtgct gctgtgtca gcaatccagt tttggattga gatgttaattc acatactatg 31800
aagtctgcct tcgtcccttt tatttgcagt gctgagaact gaaccgcgtc cagccagatg 31860
ctctactatt gagctacatt actggcccaa gtttgcctt ttaagctgca tagtttagtgg 31920
gttgcacagt tgtatatacca tccctactgc ctaattccag aaccttttg tttcaccctt 31980
aaaaaaaaacc ctcacacata ttagcagect cttcccatct ccccaccaca ctccctcctc 32040
agctctgaga tatcagtaat atatttcca tctttgagaa tttgcatggt ctggacattt 32100
tgtagaaatg aaatctgatt ctatgtgcct gttgtgtgac aaaactttta ctggggagac 32160
aaaacacata catctgttca cctcagatag ggatcccattt ataagccaaa ggatggatac 32220
caccatagtc caacattgtg aaaaaacagg ttttatgggaa gttacttaca ggaacagaaa 32280
tgactcacag acagctgctt accccagcac aggtgacago tcacaaaagc tggttgcct 32340
gaacacacag cacagcctgt aggcaagtc acaggtaga gcgtgttctt tcttagcaac 32400
tctggctcaa gcctcttcca ggcaggtat cagctcttagt agtcttctgt gcagcttcc 32460
tcctctaaga atcccttggaa ttcaagttca tgtcacgtga gaaggactct cagctctaa 32520
tacttattct ggcagaaaaag gaggcttagt aatctggtca gtttcaggaa cttcctgaat 32580
caatttgagt tgtcttcctg cgtaaggaac ttccaatagg gtggaatgtt ttaagctagg 32640
aggagattgt tacacaacat gtcaggatgt cctggctcac atgggttctt cacggcgccc 32700
atgctgcctg ggcacatcagga tcttgcctt ttcatgactg aataagttt taccacgtgg 32760
ccaccccccata gtttgcctt ctgttccctca tctggaggcc atattgggtt ggaggtttt 32820
tttgtaaatt acttccttctt cttcacagag tggccggct ttgtgtcagc cactatcatg 32880
acccccagtt gttccctactg gcttttata tcatctcatg ggtgcgtgtg atagcttcaa 32940
cctcttgagg tccttagata gccttaggctc ttcaatccctt ggtggagctg ggatcaccta 33000
agtcatggat tcttcataatgt gtgggttccac cagaactgat accacagcca aggacaccatg 33060
caacctgtga tggcaagccc agttcttta ggctgagtct gacccagagc accacatcac 33120
tctgtgttag tctccctctg agtcccttc cagggccctt accctgggtt agcagcttct 33180
gtgaaggct tgggtgtgg cccctttcc aagatcttt taggtgcctt ggtgtattca 33240
aactcatctc cacccctctg cctcattttgt gttccaaag ttctactta gagtaggccc 33300
agaagatgtg aagacaccag acatctcagg tggatgggtt gctgcctctt aatctcagca 33360
ctggcagaca aagacagttt aatttcttagt ttcaaggcca gcctgatata tagagttagt 33420
tccaggatgg ccagggttat acagagaaac cctgtgtcat gagaaatcaa ccaaccaacc 33480
aaaaatgaaa agcaaccacc accaacaacaa cagaaacaaa tcaacaacaa caacaacccc 33540
cagactcagt aaggaactcc ctggatccctt agcagtttgc cttagcataat ccaggcaat 33600
gaacaaccat gactcagaaa aacaggatga aatgtatggag gtgttctgaa cacagaccta 33660
aggttgcctt ctatgttccca catacatgaa aacatatgca cacatgcaca catggatttg 33720
tgcacatgac acataatgttccca acacacatgt acacacatgt gtgcacccat gcggacacac 33780
ttacacacat gataaagctg acctgttttgc ctggaaagct actttaattt tcagtgctg 33840
agacataacc caggccagttt attccagaat ctttgacgtg gtacctggtc attggatttt 33900
tcttaaatcc cccaggaaat tcttctctgc gtgcagggtt aaaaccacta caccactagg 33960
gtccatgctg gcctaatca tagatgcgtt ttctcttacc aagcctttgg aactcagtca 34020

-continued

agagcttagt caagaattgg tctcaggagc cgggtggtgg cggcgacgc ctttcttaa 34080
tcccaagact cgggaggcag aggcaggtga atttctgagt tcaaggccag cctggctac 34140
aaagtgaaaa aaaaaaaaaa ttggtctcag gttggcttg tatttctat agaattctgg 34200
tcaagttctt ccctctctt gttttagcaa aacaggcctc acagagatgg agagatgact 34260
cagcaattaa gagccctggc tgcttcca gaggatccaa gttcaattcc cagtaccaca 34320
tggcagctca gagccaaactc ttaactccag gtccaggaga tctagtgcc tcttctggct 34380
tccataggca ctgcacacac tgcacactat gtgatataca aagatataatg tgggcaaaac 34440
acccatacac ataaaataca aataaatttt aagcaaagtt tgggttatgt cagtggtttg 34500
cctacaaggg tggggagtga gaggtgaggg ttgggtgtga gcatgtattt aaagtcaact 34560
gggagcattt tcagaaatcc aagtggcag ctcaactcag aaagttgaga gcaacctgtg 34620
gtttccaacg ctgccagcgt gaattatcct ctctcggtag gaagcggtca gtggaatctt 34680
gtgtcgaagt gaatgattt tttttttaaa caagtaactg cagcactaat tattgtggca 34740
tttcctcaaa ttccccatctg tgccgcttac gctgtttgc actcatgcca gcaacaacta 34800
aaccccttgc tttagcaaga gcctcaccag taggatatac tgggtttttt ctcagggttt 34860
cgctagtctg atcagtgaga ggctcagccc agtagcattt ctctgggtgt ctccctctg 34920
agaaggcgtg agtgtttct cagggttaag atctgtttgt tttccctccca taaaccattt 34980
tggtttttt tggtttttc cttctctgaa cctagagtcc catgcacact gggcaagctc 35040
taaaagtgaa ctgcgtccct tagtcttca tttgtttttt tattatgaga caaggtcaaa 35100
atctgtcagg tggcccttgg gttacgattc tccctcgatc ctggccctga cttctatgt 35160
ctttcttgc ttccctaaccc gcttcctat tgcagatatt tccccagct cattgcctat 35220
ttttaatgc acgcgcgcgt gtagtgcgtgt gtgtgtgtc gtgtctgtc gtctctggga 35280
tgggacttgg ggctttgtc tctccctcg agctgcacag catggattgc tgcgtttttt 35340
tgtgttccat gcagtgatc cccgtgtat caatgcattt atccccctcc ctccaaacgct 35400
tctgggtttt tcaatcctcg tttgtgtatc gtctctctca gtcatttca gaaatgtatcg 35460
agtctcctgt gctgtttata gtttctacc cgtatttgt ttgttagtta gtcttcatcc 35520
tgtgttcagg gtcctacgaa atgtcacttc cccgtgaggcg gagacttgc tgcgtttta 35580
gagacaagcc cttttggctg tgaggaggatc agtcacatca catagatggc ctcagctgg 35640
gctagtgggg acctgggata gtcttttagat gtgtattcca cacaagaaga gcatataaaag 35700
cgccccctt ggccagcaag caccactaag aatggccctc agcctggagg gctttctgccc 35760
cacagcatcc ttagcatctc atctgtctag ctcactgcta tgcgtctac cctctgagat 35820
ttggccggga aacctctgat gagcggatga aagacaagcg tcagccatga aaacatacat 35880
acatgtgaca tctggagtga gcaggatgtt tttatataatt taggaacaca catatatgt 35940
gcaacatcta aagaccataa atttgcatttga gaacaggaa ggggatacat gagaagggtc 36000
agagggaaaga aagtgaaggaa agaagatgtat gtaattgttag tataattggg gggggataaa 36060
aagccccaaa catcaaagat gaacaagagc acaaattctg atttccagcc aacaggacca 36120
cagaggccag atgggtgcag tgggtactg cacccctgtg ccgggtgtcta atcatttctt 36180
tctttctttt tcttttcttt tcttttcttt tctttttttt tttttttagaa 36240
cagggtttac ctgtgttagcc ctggctgtcc tggaaactcat tctgttagacc aggctggcc 36300

-continued

cgaactcaga aattcgctg cctctgcctc ccaagtgctg ggattaaagg cgtgcgccac 36360
catacccgcc cttagtcattt ctcccagtagt ttgtggtcat tctctcctgc acaagggtgct 36420
cggtgatccc tggttccgct gtgggtaacg gttagcacag acaggctgag tggggaccct 36480
cagtagtact ctgtttatg acacctttaa ggtatgtgtt catcttgcgt accttcagt 36540
ttgctttctt ctcccctt agcattcctt gaagaaaaga ggaacccggt ctctggggaa 36600
gacagataag aagcctactg cacaggtagg acctctgtac ccatcacact ccctgcagt 36660
ggtagtttg atgcccatac caacttcaag tgtcctgagt agtccccag aggtccctat 36720
cccttcccaa tgtctctcctt gccttccttc ccctatgcat tttttggct tccccattta 36780
gttcctccaa gccactccg accccatct tacacactcc tggacccaca agttgataac 36840
ccaacatagg gacggataga cccaaaggga gtcggccctt ccgcagctt actgccacc 36900
gactgagaca aatcagcttt ccagtcgtt tccatgcctt tccatgttctt gtgggtttgt 36960
tggaagaact ggataaaaat gtcatgttgtt gcacaggcca gaatcctgtt agcagccaag 37020
gcacccctt aacgggttcc catttagcag tcatcatatg acaacaatgt ttctccac 37080
ttttttttt gttttttttt ttgtttttt cccgagacag tgggtctctg tatagtctcc 37140
gctgtcttag agtcactttt gttagccaag ctggcctcga actcagaaat ccacctgcct 37200
ctgcctccaa agtgctggga tttaaggcat ggcacccacc gcccggctt gcccacttt 37260
agagatgggg caggaatggc catagtcaca tggctgaagg tcacctgctc ccagactcc 37320
agattgtagg tagtgtgtct tgggtctca atattgtatg ccctgctctc tatectgtgg 37380
cctccaggcc ttgagagcca agctgattca cccctgtgtt ctccgacagt cagggcaccc 37440
cagaatcctg tggctgggtc caagctgaca ctaaccttcc ttgttggctc ctgtctccgt 37500
gtccattggg ctgcaggagg atagtcaga cctgaagtgc cagctgcatt ttgcaaaagg 37560
ggagtcggcc ctcatgtcga agaagctcac caagttggct aaggagaacg acagcatgaa 37620
ggaggagctg ctcaagtaca gategctcta tggggacctg gatgcagccc tggcggcaga 37680
ggagctggcg gatgctccgc actccctgta gactgagctg aagggtgcacc tgaagctgg 37740
ggaggaggag gccaacctgc tgagccggcg catagtgag ctggaggtgg agaaccgtgg 37800
cctgcgagcc gagatggacg acatgaagga ccacgggggtt ggcgggggtc ccgaggccag 37860
gctggccctt tcttctctgg ttgttggatgtt cggggagacg ctggccaggt tgccggccca 37920
cctgcagttc gtggaaagagg aggctgagct gctgaggcgc tccatgttgc agctggagga 37980
ccagaacaag ttgctgtga acggatggc caaataccgc tggagacacg agctggacgt 38040
gacgctgtcg gaggacagct gctccgtctt cagcgagccc tcggcaggagg agctggcagc 38100
cgccaaagctg cagatcgccg agctcagccg caaggtaag aagctgcagt atgagaaccg 38160
cgtgctcctc tccaaatctgc agcgctgtga cctggccctt tgccagacgca cacggccat 38220
gctggagacg gacgctgagg ctggggactc tgccagtcgatc gtgcctgccc ctctgggtga 38280
gacgctggag cccaccccg cccggctgtg cagggccggtaa gggccggagg cgctggccgg 38340
cctacggggag caggccgctt ttgttgcacaa ggccatcgac tccatgggtgg ctgtatgcca 38400
tggcttctca gtggccctcc gcctgtgcgtt ggacaatgag tggctgtact tgccactgca 38460
cgaggccct gacaacacgacg agggcccccag ggtatgccaag ctcatccacg ccatcctgg 38520
cgccgctgagt ttgttgcacaa agggactgaa cgccttcacc cgcaaggcag atgtggccctt 38580

-continued

ggggagctct ggcaaggagc agcctgagcc ctccctgtc ctgcctgcct tgggctccca 38640
gggcctgct aaggagatca tgctgtccaa agaccttggc tctgacttcc aggtaaagatg 38700
ctacatgttc tgaaccaggc acatgacaga aggacacaaa gccccaggta cacaacttca 38760
caactgccct gcagggagca gtccctggta tcaccttagaa cataaccagc cctaaggcag 38820
ccatggttca tctccctgtt ctctgggtt attctgtttt gagctttata taaaattctat 38880
agactaaaca cagcagtgcc tggctgtgtt ctctcagcca actatttgct gttacaggta 38940
ggaaaaacaa ttatatttattc tcacctaact aggaataagc ctgcccactc agcaaactgc 39000
ctggtatgcg tttactatgg atagggatta attctgaggt cccatcttt tgctagactt 39060
cccgtttttc tccctttccc tctccctttc cagtgcgtgg gacacaacgc aggacctcat 39120
acatgttaagg caagaacctt accattgago cacatccctg gccccttctt tttaaaaaca 39180
ttttttagcc aggccgtggtg gcacacgcct ttaatccag tacttgggag gcagaaccag 39240
gcagatttct gagttcgagg ccgccttgtt ctacagatgt agttccagga cagccaggc 39300
tatacagaga aaccctgtct taaaaaacgg gaaaaaaaaag aaaaagaaaa aaaaaaaaaa 39360
aaagaaagtt tgcatatgtg tgcgcagggt tgcgtgtgtg tgcgtgtgtg tgcgtgtgt 39420
tgcgtaccca cagaaggcag agcaaagtgc tggagcccc agagctagag ttgaagggtg 39480
ttgtgagctc ctgtgattagg gtgcgtagaag ctgaactcggt gtcctttaa agagcattaa 39540
tgtatcttaa ctgttgcggcc atttctctag ggttagcccccc tccttttcc atctttttt 39600
taactaaaat tattacacgt atttgggtgg gttcaggggg agccgttgcgtt cacatgtgtt 39660
ggtcacagga gtcagtctca gatggctctg gggatcagct caggcttggt agcaagtacc 39720
tttacctgggt cagctcgctt gccagatccc cacatcccad ttcttttgg tttttcggag 39780
gcagggtttc actgtgttagt cctgcgtgtcc tggaaactcac tctgttaggcc aagctaacc 39840
tgaactcata tagagatctg cctgcctctg ccccccccad cccccacaaa tgctgggatt 39900
aaagggtgtt gtcatcaccc cccaccacca gtccataata ttaaatttaa aaaaatattt 39960
taaactttt gatcaacttt tgagaagttt agccttgata cccatgacag gctacccaga 40020
aaaactaaaa cgaaaaagac acttggtaa aaatttaaa tgcattagtg caaaagcaaa 40080
acgttaaacat ttgacgcaaa tataaaaata aacgaacacc tcgtgtgcct acctaccact 40140
tcctgcggat gtctctgtac accgtcttga ctgggtgacgt tcagtgcagc agccatgtgg 40200
ggatgtgcgg cgatggtaaa ggatcagect tagttggcgg cacaggttct atccccagta 40260
cagaacagaa acaacccaaa actcccaaga caaatttaa aaattaaattt aacgaagcat 40320
aggattgagc tgcagttctc tgcgtgttagt tacacatcgat agatgcttgc ggctggctgc 40380
gtgctgggag cggccctct gatccttccc acacagccctt tcctggccctt gggcttcatg 40440
ttatggctac tgcgtgtggg cagtggtgtt cctcacaccc tgcattccctt gtccgcaggc 40500
tgggataacct tgcgtgtggt cccccccctt cgtgcctaaac acgggatggg tgggtatctg 40560
cagaaccagg gagggatagg agccccactc gaaaggtccc accgccttgc gagggggaaat 40620
taattagtca atctgtcaag cctgtgaatg tgacatgaga aatggttttt atcaaccagg 40680
aacttgcctt agatctgaa ggggtcaga tcattgagag aaggaacacg gtaggtcga 40740
ggaacaggac agagttgtga ggctggctg cagagagcgt ggaggtgggtt ggttttcatc 40800
ctgagagaat ggcgggagat tgctgtgtgtt ggccctgcac tgcgtggac cggccgggtt 40860

-continued

gtggcttcag ctgtgctgct gtggcttcag atcgcttctt gaaaggggag gcccgtttg 40920
agtctccctc ttccctgaagt aagagtgagc agggaccaca cgggtcaccc acagggtat 40980
agaattgttt cagagggctt tgcgtaagag ccattcacaa aattaatgtt tccacttgcc 41040
ccactgctcg ttatgcaata agcaaatgag ttgtggcctc gcaccctggg aaggccaatt 41100
caccaggctc ccctctcatg tggacaaagc ccattatatg gcagaggctt ggtggggacc 41160
tttggcact tccttggcc tctccctatg gatctttctt gcttttagcc acctgacttc 41220
agagacctgc ttgagtgaaa gcccaggatc cgagaggctt tccgttccgg ggacttggag 41280
tccaaaggctg acccttagtcg gaacctcagg ccctaccggc ctgaagataa cgattcttat 41340
gcctctgagg tgggtctagg cctaagcagg ctgttgggtt gaagccgatt gcctcctgaa 41400
agcaggcagg gtggcccttg gaggaggatg gtataggttt taagcattta aacttctgac 41460
agccaaggca aaaaaagaaa gaaagaaaga aagagagaga gagagagaga gagagagaga 41520
gagagagaga gagagagaga gaaagagaaa agggaaattt aatcaatccc gtggtttagt 41580
tagaggaagc ttagctggat atggacatg atatcaaact gtaatcccag tgtcaggacg 41640
ctgaggcagg ctgaatgctg agctggaggc catctgaaa tgactaactt atgttagcaga 41700
gaagcaatgg cctctccatt taggctgttt gagttcatgc tggactttc tgggtgctgg 41760
ctgtgagatc tcgggtacgt catttcactc tcgaaaattt tcagggttgc cacagacaag 41820
tacagagcct cataggattt ttggcctggc ctccacgggt tgaggatgag gtcgggtgtcc 41880
gtgacaagtc acatggactc agggatcctg agtgcgttgt cactctgctg gtctgggtgt 41940
gtgacccctcg acatgtccac tccccatggc cccggctttcc caccacaaca gttataatcc 42000
ctagagttgc taagattctc tggatcttgc gaagttccctt gtcctcctgaa gcctggggaaa 42060
ctccctggaa ctgaagggtc caactgccta gtgaggtggt ctggatgtcc tcataaccat 42120
ggcgaaggag gaagaggaag agcgttgtctt ggggctggag agatggctca gtggtaaga 42180
gcacccactg ctcttccaaa ggtcggttgt tcaaattttca gcaaccacat ggtggctcac 42240
aaccatctgt aatgggatct gatccctct tctgggttgt ctgaagacag ctacagcgta 42300
cttacatata ataaataaat cttaaaaag aaaaaagaaa agaaaagcac tgtctggaca 42360
aggagctaag ctgttagtgct gtgttagcattt agctttgatc cccagtcagg agatggatag 42420
acagagagaa ggggaggaggag agagaaagaa gaggaggagg aggaagaaga ggaagaagag 42480
gaggacgatg actacgttgc agcaacactgatc gatgcgttgc tctcggtttc tgagaagaaa 42540
ataggcttgtt ttgttttgtt tggttcattt acacccgttgc tccgtttccct tcattcttac 42600
cgattagctt tggatcttgc tggatcttgc tggatcttgc tggatcttgc tggatcttgc 42660
gcccctgact tctatataag aggtttttt tttttgtttc tttgttgc tggatcttgc 42720
gtaattatataat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 42780
ggattacagg cacatgacag tttggccaga tctgcagaga ttttttataac agaacaggtt 42840
tacaagacac gcctcgatc actgcagaga cggccaaattt atattattgg ccacatcccc 42900
tttcacagca ggaacctgtt atgcgttgttccatgc tcttctgttgc tggatcttgc 42960
tggatcttgc tggatcttgc tggatcttgc tggatcttgc tggatcttgc tggatcttgc 43020
ttactgtgtt gtcaggctg gcctcagact cacaggttcc ggtgatccgc ctgttgc 43080
cccacttcat ctcatctaaa aaatgttgc tttttttttt tttttttttt tttttttttt 43140

-continued

tggggctgca gctcagtat gaggcactt ctcagcacgt gtgaggectt gtgcccacac 43200
aaggcagggt ctgttatgt aaccacgac acgtgcacac acatggactc acatgcacac 43260
acaagcagac acgtgtgtac acacacatgc actgcacatg gacacaccca tgtaccacac 43320
atgcatacac attcatacat gcctgcacac atatgcacac atacatgcat gtatttagtca 43380
gggttctcta gagtcacaga acttatggat agtctctaga tagtaaaggaa atttattgtat 43440
gacttacagt cggcagccca atccccaca atggttcagt cgagctgtg aatggaagtc 43500
caaggatcta gcagttactc agtctcaegc agcaaggcagg cgaaggagca agagcaagag 43560
ctagactccc ttcttccaat gtccctatat tgcgtccagc agaagggtta gcccagatta 43620
aagggtgttc ccaccacacc ttaatccca gatgaaaggc gtagccaga ttaaagggt 43680
gttccttaaa ctccgagatt caatcttctg gaatccatag ccactatggc tcaagatctt 43740
caaaccaaga tctagataag gatctccaag cctccagata agggtcactg gtgagccccc 43800
caattctgga ttgttagttca ttccaaatat agtcaagttg acaaccaggaa atagccacta 43860
caatgcatta cacatggata cacacatgc ccacacatgc atgcacatac atgcacatac 43920
acacacacgg ggaagggcag agatggagaa aggaaggaga atgggtgcct atggctttt 43980
acaaactgaa cacttcttc ccctttggtg ctggggcag ggtcttgcta tgtaacccac 44040
gctggcctca agtcatgac ttccctgttt gtaagtaaca tctagaatca taagaagact 44100
ctagacatct gtgttcccta ctgtctgtct gtctgtctgt ctgtctatct atctatctat 44160
ctatctatct atctatctat ctatctatct tgcgtgtggg tatgtgcaca tggtatataa 44220
gtcagccggac agtttatagg agttgggtct ctcctccac catgtgggtc ccagggatca 44280
aacccaggcgc ctgacgcttg gctgcaagca cctttatctg ctgagccatc ttgacagccc 44340
aagaacaaat taacttcatt tagctctaaa tgctttaaa acacaatttc ccccaaaaa 44400
tggagaattt ggtagcttt tttaaaatga tcttactcca tagttactga aatccacaaa 44460
taatcacaat taaacttcgt ggcagcatta ttacatgaag aaacaaagaa aagcaagatc 44520
attctgttagt taaccacacc gacgttcacc gtgttctgt tacgtaaaca cgctcaaaga 44580
cagcatctgt gacggAACat ctggatggaa cttaacacat acagatccac taaacacatc 44640
tgaaacctca ttatctttt aaactgcgtc taaatttgc attatagaca gtgccttagc 44700
ttggggatgg caagccacca tagcccccac tacctgggg agctgaagta gggggatacc 44760
ttgagctcgag gatgggtccca ccagccctgtg agactctaca tcagaaataa gtaatcaatt 44820
acatttaaatc agttaatttg atcaattaaa gagactatga acaacgttgt gacttacatt 44880
gagttgcac aggatttttcc ttctgtcag ctacgtttgc aagagagaaa tggtaaaaatt 44940
atattggaa gatttgacac gtactgtttt gtgttaattgt cctgtatgacc cgctaggcaa 45000
ggacagtgtt gtgcctcgta gaggtgactc ccattccagggtt ttctgttgac tccttactgc 45060
ttctgttctt tgaccacagc ctgtgtcttc tccttctcc ttgtctgtcc cctccacccct 45120
acctccctcc tgcagatcaa ggatcttgc ctgggtctgg cccggccca cgacagccctc 45180
cggggcttgc aagagcagct gtcccaaggag cggcagctcc ggaaggagga ggctgacagc 45240
ttcaaccaga aatgggtcca ggtgggtttg taccacaggt ttgtgtttt tggtaaaaatt 45300
ctgactatag ctgtgtcgag aagggaaagggt aaaggggccca tgggatagct ctgggttgc 45360
acacttgacc agcgcgtgag aagctctggg ttctgttccc agcataggaa gaaatcaaacc 45420

-continued

actactaaaa agaaggggac aacactgggg aaacccaago ttccccaccc ctgcgcctcg 45480
ccagtttgc gttgtcagca gctacactga gctggagcca atcacatggt aatagtttc 45540
tgggtgacgg tgagcctgct taatatttct gggtcttgct tggttgtgtt tggttgagaa 45600
agggttctc cgttagcccc tggtatcca cttgatatgt tgactgggtt gatTTtaact 45660
cacagaatac ctccctgttc tgcctccgg gtgcgtggat taaagacata cagcaccacg 45720
cctgtctctg tttcatgtct ttttttaaaa aaatttttt taaagaattt ttactttaa 45780
ttacatgttag atgtgtatgt aaatgtgcat gcctgttagt gtatgtgtt atgtatgt 45840
atatatataat atatatgtcat gtttgttaa gtgtatgtgc atagatgcat gtctgttag 45900
gtgtctgcag aaaccagaag tgcagggtt ccttgatct gcctaattgtt ggtgcttaga 45960
attgaatccg ggtccttagta ctcttaacca ctgagccatt tctccagctc ttgtttcatc 46020
tcttggcctc agtttctgt ctgtgaaatg gtgcctgtt caaacctggc tggatcag 46080
atctataataa gcattttttt cttaaaaaaa aaaagattca ttttattttt tatatatgt 46140
tacactgttag cggtcttcag acacaccaga agagggcato agatcttattt acagatggtg 46200
gtgagccacc gtgtgggtgc tgggaattga acacaggacc tctggaaagag cagtcaagtgc 46260
ttttaaccgc tgagccatct ctccagcccc tctaataaca tttctttct tttctttct 46320
tttctttct tttctttct tttctttct tttctttct tttctttct 46380
tttctttct tttctttct tttctttct tttctttct tttctttct 46440
ctcttcttgc gagacagggt ttctctgtgt agccctggct gtctggaaac tcactttgt 46500
gaccaggctg gcctcgaaact cagaagtccc cctgcctctg cctcctgtt gctggattt 46560
aagggtgtatg ccaccatacc cggctataac atttcttaaa ccttaatgtt gaggctggtg 46620
agatgggtaa agacacttgc caccaaggct cacaacttgg gtttgcattt caggaccac 46680
acagggagag agagctaacc tgccctgagtc ttctgaccag cacacacatg ccacggcagc 46740
catgeccacc catacacata caaaacagta agagagtgtt aaaaattttt agcacctcat 46800
tggacggccc gtggaaacca ccgtgaacac atgatagctt gcaggccctca tactcctccc 46860
ttataagata ccctcaactt accgatttgtt ggtgcgtgtt aggatgaact gaagaggcag 46920
gcagactgca gacaggagca cagacgtgtc cagtcgtgtt ttgctcagag ctgtacatcc 46980
caggctccct tagctgtgaa ccagtgttac ttactccctg ggggtctct tgggtctcag 47040
ccttctttaa cccagcatgc cttgtctgc acccaccctc tctctgtccgtt cccactcac 47100
cagtgtgagc acgtggggaa cacctgtgc tctgtgggtt ctcagagttt ggtgtccaga 47160
gatcggtctt caggctgttt gtcagggtt ttgagatggg atccctccctt ggcacctggg 47220
ctcacctatg agactacgtt gggcggctga gtttgcattt gtgggtctt gagagtgtt 47280
tctggtctt tgcttgcgtt agaggcactt ttagtgcactt ctctctccag cccactgtgt 47340
ctgcttgcgtt gagttctctt tccaggata caatggctg ctggtagtag aagccctggc 47400
ggggggccccc cagtaggcac ttgcctcatg tggccatcatg tacccttgc tgggtctgca 47460
atcaacttggg gcctcccttgc tcaatgcacag ctgaaggaaag accagcagag 47520
ggcgctgtgtg agacgggagt ttgagctgca gagtcgtgtc ctccagggc gactggagca 47580
gaagttctgg agccaagaga agaacatctt ggtgcaggag tcccagcagt tcaagcaca 47640
ctttctgtgtc ctcttgcattt agctccgggtt gttcctgtttt cgcgtggccgc agggcaagg 47700

-continued

tctgcccagc gaagaggatg acttccttga ggtatggctg agcatagcct tagcgccggg 47760
ctggaagaat gggagcgtgg cacctgagtt ctgcctcttgc tccctaacag tacctggtt 47820
ggagtaggtc acatgctatc tggactgctt gcccgggg tctcccggtt tggtaatgt 47880
ggtttctcaa accaggtgaa cagcatgaag gaaactgtacc tgctgatgaa ggaagaggag 47940
atgaacgccc agcaactcgga taacaaggcc tgcacagggg agagctggac ccagaacacg 48000
gtgtgtttaa ccccttcggg tcctaccta actgtgtgta actcggtcat taatgagcgc 48060
ttcctgctat cttctcagcc ctgacaggac acagggtctg tggatgttta ttgcttggat 48120
tctggcccac cagtttctca gtgctagggg gacatgccag ggttagtgcaca gaaggccaat 48180
gtcacaagt a gggaaagcag ttcccccgtt tgttttatag gaaacagtgt tagagaatgg 48240
gggtcccggt tccgtcggtt tccatggat aataggggt tgccttccatc agaataggaa 48300
aaaaaggcca tattgaaccc agaggacagc atgtgtgag gtgtggcaag atgtgtgg 48360
gtgctgctgg ccagtggac attgtggag ggtgaggata taccaggcca gggacatacg 48420
aagagagacc ttaagtaaa actgtcaat caagccacag gttggagagg gcccgtgg 48480
ccccccctag gggtctatag tgagcttagag tgaggccgag cagtgtgaaa ggcaccagga 48540
aagagggtttg ggtccagtc tggccatt tgaggacat cagtcagaga ggagagacct 48600
gcagecttggaa gccccctaag cctaggtgag gttgtttctc tgcgtttggg cagatgtct 48660
gctcctggaa tagacttcat tgcgtatgag actgtgtgtc cagatgtgtc atttggac 48720
ggccttcagc cccatttac tggtgagagg acctgaggag agtctctgg agctggtagg 48780
gccccagctg gactcctggg tgagtctgtc atgcccggc tgcctttct gccttcataa 48840
agttcctgac caggcattac tctgcccggca gctaatgag tgcataaga ccctggccg 48900
catgaagggtc accctgaagg agctgtgtc gctgctccag gacgagcgtc ggggtctgac 48960
tgaacttcag cagcagttcg caaaggccaa gcccacccgtt gagacagacgt gtgcagatg 49020
caagggccac gctcgcagg taagcctccc tgcttgtag ggacgggtccc ttcccttttc 49080
ccggcccgac tctagacagg caggtctgg aggaggatgc taaggatgca cagggtcccc 49140
tgcacccatc cgtgtgtcccc agagagattt acaggaatgt aggctgcact gtggactcac 49200
agaacgcttc cacgcccagcc tgcgcgtcat atcctgcattt gctcctaata gaccggacta 49260
gggtgttgag agggtaaga aaagacagac agacacaagg atacacatgt agctggatc 49320
aggcgatctg gactctctat ggagctctgg aaactcagcc tatttaccat gaagttggtg 49380
gtcattaatg cagataaccc ggggtgggg gattatgggt taccacatgt tcaggggaca 49440
agctcagcaa atctgagcag gagegctctg taggggagct gtctccaggc tgcgtatgtt 49500
ggagggggct gctgtggccg acatcttcat acactaataa catgcccac acacagagg 49560
aggcttgcctt gttcccttgc ggtctccacc ctgggtggac aaatggctt gccatgtctc 49620
cctgggtccc acggcgcgc tggatctgtc acgtgactgt acccattca agagtacaca 49680
taagctcagg acctgactca cgaggccgtc ctctgccttca acacatgtac ttggtaacag 49740
gctgcacact cctgtctggg ctctgtggc tgcacccctt atttgttttc ttctttctt 49800
ctaaggactc ccagccccgtc ctactggctt ctaatgtgtt cctcataactc aaattgggtc 49860
tccccctcaca gccacccagc ctgactaaca ggcgtctcac cagcatcccc agggggctgg 49920
catctgctgt gtgcctgggg gcaagttgtt tccccactct gatctgacaa ataggaagg 49980

-continued

agaaatgctg tctctgttgc tctggactgg gcccctggcc gcctccagca catcttgct 50040
tcgggggtgcc ttcccttgcac tctgtggcca tgggagacag gcatgctgaa ggtctctgtc 50100
ggccagatgg ccaaggcaggc ccagaggagt taggtttctc tccttaggtat taagtagcag 50160
gcttgcgttt gccccccagg ctctgtaccc tcttggtccc atgggtgcat ttccaccctc 50220
gcctgggtgtg tggagactcc ttgagttcc ctgagccctg cgcttccctc caccctgtgg 50280
gtatccatct tcttccagcg ttccccgtg cccctcaactg ctccccaggc ccagccgtcc 50340
caggagctct ggctctcgga ccagcctcat aaggatcago gttcttagatc catgggcagt 50400
gagtttttagt gctgtatgtt tcttcctgc ctccccatga tgaggtcgag gctgagaaaa 50460
atatgctgga ccattccctgc gccagcgtat tatgtgactg tgtgtacatc tccctgtgca 50520
gggcctctgg gacaagggtga ggcccaagtt tcaaggggta gcagtggggataaaagaatg 50580
tcttccaccc gtggaaagctg gtgtcctgg taagaactga taaatgtttt gttatcagta 50640
gcggggggctc tccctttgt agcaggatgt ggaatgctta tctagccct tggagccac 50700
gtgagggaca ggaagtgggtg ctttgcttac agaggcttag gaagagaaca ttgggttcegg 50760
ggtcagtctt gtatgttagaa aggggtccat agcacagaga gtccctgaaag acctccccag 50820
tatgggtttt gagccagtgt cccctgccac tatccagcta gctgggtctt cttccctgcag 50880
atggagctga aggctggaa gggtgccagt gagaggcccg ggctgactg gaaggctgca 50940
ctgcagagag agcggggagga gcagcaacac ctccctggcag agtccctacag cgccgtcatg 51000
gagctgacga ggcagctgca gctgagcggag cgccactgga gccaggagaa gctgcagctg 51060
gtggagcggc tgcaggggaga aaagcagcag gtggagcggc aggtgaagga gctgcagaac 51120
cgccctcagtc aggtgaggag gctctgggtc agtatactg ggagcccagg gcacaggggaa 51180
agccgccaga gtcggggccct tccctccca gcacaccaca tggcagcctc ggtgatcgca 51240
tcttacatct tagcagttgc aataaggaag tgatcttgc tcacgagttc tcactgac 51300
ggtgtcgate ttgtgcccattt tggtggggcc tgcgtgagacc caggaccata aagggaaagag 51360
ccaagagtgg tttccatagg aaagcggagag tcctcagacc agaatgggaa accaagaccc 51420
aagtattaca aaatccttat tccgtccggc tccctccctt tgctctgcag gtatacatct 51480
gtggagagat ctctgtctgg tttggttttc actaatttagt gtcattgtctt ccgaggagag 51540
ccttgcgtcggtttaactgct gtttgactcc cactcattcc gctggcggcag agtccaggct 51600
ccatgtttga ttggagcggcag gtgacgtcac tcagttggta cagcactgcc ctagcacgca 51660
tgaaggccctg ggctggattc ccagcacccac aggaacaatg tggcggcagct catggctgtc 51720
agttttgggt aggtaaaggc agaaaggta gaggttcaag gtcacccatc tcagctactt 51780
agttagttt aagccagccct ggttacatg aaaccttgc tcaaaaaacc aaaccaaaca 51840
aacacaccag cagaaaggcag ccctctactg tcctgtgcac aggtgtatat gtgaacagta 51900
tttacaatgc tgcgttcaca ggctctcaca tggatagaga gtgagagaaa gagaacttca 51960
ggggggcttc cccactgttc tcgtgtctgt ctgtctgtct gtgtcgtatc cgccgactcg 52020
ctttgcgttct ttcttgcac tctccatgtac tctccagttt ttggactcac ctaatcccc 52080
tgagagccac catctttgtc ttgtcagecc gccatccaa accctctcag tcctgtaaagg 52140
ccactccca gaagagaaga acgagccaaa atgtcctgtt aatgacagga cagaacatg 52200
ccagctccctc acatctgttg gatttttttt ttttaaagat ttacttgcgtt ttatgttagat 52260

-continued

gcgtgttttcttgcttact tacatgtaca cacaccatat tcctgttcat attctctggc 52320
tagtgtcaga tttctcagtt atagatggtt gtgagctgcc gtgtgggtcc tgggaactga 52380
gcctgagtcc tctggaaaca agtgctttta actgctgago catctctgtca gcccataacat 52440
ctgatagatt cttttttttt tttttttgtt tttttttttt tttagagacaag gtttctctgt 52500
gtagctctaa gtgtccctgaa atttgtgtg tgtagaccgg gctgacttca aactcacaga 52560
gatcctccctg cctctgecctc ccaaattgtg ggattgttagg catgcaccac cactgccccaa 52620
cttttaatata ttttttttaa tgatttgaa atcattaacc aaatcaatgc cctgctgaat 52680
gaggtgttag cacaccgctg ggacctgcca gctcggtgg gctcgcttctt gtgtttaatt 52740
tcccaggcgc gccccacat ggaatttgct gacttcctag ctggccttgc gcaacctttt 52800
gtatgtgttatt tggtttccctg cattcaagcc caagctgcat cctgttatttc ttatttttttt 52860
cctccctttt ccottcccccc tgccttttctt ttctttttt gtttgggg agactacata 52920
tctcaatatt tggcccttggac tagectggca tgctcagttc tcctgagcta ggatgacaga 52980
ggtcagcccc aacatctgtta taagctgtgc tcttagatgaa accagcttgg tcaagcaacc 53040
agcagagcaa cagataggaa gagtatcttt agtcagtgtt ctccccacg aagagtacc 53100
ttgaccatcg cagctttaa aaggaaacat ttaatcgggg ctgggtggctt gttccagagg 53160
tttcctccct ttgcgtcatg ttgagcagct tggtggcctg caggcagact tggtgtgg 53220
gaggtggctg agagttctac atccagatcc aaaggcagca cgaagagaga gtgacaatgg 53280
gcctggcttgc agcatttgcg agttcaaatg ccccgcccc cagtgcacaca cttectccaa 53340
tgaggccacg cttccataata gtccactcc ctataagctt atgggtctca tttttttttt 53400
tttaaatatt tgcctttctt tttttttttt taattttata tatgtaaatgta taagtacact 53460
gtatgtgtct tcagacactc cagaagaggg catcagattt cattatggat ggttggagc 53520
taccatgtgg ttgctggat ttgactctgg gacttttggaa agagcagtcc gtgttcttaa 53580
ccactgagcc atctcgccag cccctatggg ttcattttt attcaaaacta ccacaaagg 53640
gctccatcg ctaatcaaat cccctctgtat ctctggctt cctctacaca aggaggatgg 53700
ctgttcaatc ttcttatctc tcaggacagc caagggata aaagagctt ccacaagaat 53760
ggtgcggtgc tccttaggtcc cccccctccc cccactgggc ctggccttac atcttcatct 53820
gtgattgtgtc ttcatctatgc ctgagatgac ttttggctt gcccctccac ccccttctgt 53880
tttgcttac attgggttcc aagtaagcag agcttggat ggggggttcag gtgcaaaacag 53940
gtatgtgtcttgc tgcctgtgtg ttgtttagga cttcattgc cctctggaa atggaaaggca 54000
aaggctactt gctttttcag aactgaaagg gtttgggtg gggggagcct tactaacact 54060
ggggaaagctg gggggggcgaa ggggtggggaa cttagggcag gtcgtccctg tgggagcga 54120
gtgggtatgc ggcggccgtc atctcagctg ctgttgcacat tgctccacat ctggccctc 54180
cttctgttgc ctatcttc tcacatgtca ctggaaata atctcagttg tactggctcc 54240
tatgtctgtct atgaggacta attgagatgtca tgagcacatg caaatggcc ccctggaaatt 54300
gtggtactcc agggaaacacc tcaggactca gcagggttgcg agcatggtcc ccgagtgaaa 54360
tagctactgc ttgtctgcgg cccttagctgt tctgtgaact gccgtgcaga gtaatgtac 54420
aacagctctt gttactctca tcagccaggat cagcaagatg cctgccagta gctgtttgt 54480
ttactggttt ccaaattccac ctgtgcggg gctcagaggt actatctatt gagaagccaa 54540

-continued

tgccggacat atgttgtgg tgggtcatgt ggcttataga attgagactg gggagtggcc 54600
ctgctcagag taaccctgcc cgccagcata cacactttc tcttctca tgcaaggaa 54660
ctgcccggcc ctgggtcctg aagcacatcg acatggagaa gcaagacaac agctggaaag 54720
aggtagagac atccctgcta cccacatctt attggcccg actgtaccag aaaattcctt 54780
gcttgtgaga aatgtaacag agggatgcta actctgtccc agaagccaca ccctcaaac 54840
accctgggct cccgagcccc agagtcacta ccccatgact gctgtccatg aatgtggact 54900
tgggaagggt tctctgtgt ggaagaaacc ccaggccagt tcagtcata ctcaacagaa 54960
agcaggcctt ctaagtagct gtgtggtctt gggtagtca ctctccatct cggggcccc 55020
tttcctcatt gcccctggct catctgagt gggtcgggtgg ggtgacactg gaggaaattca 55080
cacataaggt cagggggact gttaatagaa cacagagctg aggaacaaag ggttacataca 55140
ggtccctcat gtgtccctc ctcatgatgg cgtgaagaaa tggcacccca ctcaggatat 55200
ggggctacaa aacgcattttc tggagaactg aagcagagag gattttgggg ggctcacagg 55260
gctagtgtatc cgcttctgg tactcatggg ccccgaggcc tgcgtctcta gacccagat 55320
ggggaaatct cacaatcaact gtcatggcca tcaactgtac agtggctgtc atggaaaaag 55380
ttctgtctca agtggccctt ctttctctgg tcccttttg cactgcagat atctgagctt 55440
ggctccatt gccattgtaa gggggtgaga ggtaacggcc cagacagtta gattgtact 55500
gacggcgtct tacccctccag aggagacaac tcacgtgtgg agtctgggg tcagttccca 55560
cgtgggtctg atgtcatcac ccatcacaca ctggtcacca gaactgtatc tcttgcctct 55620
tcctttgaca ggcacgaatc gagaagaccc atgacaaggaa ggggtctctt gaagctgagc 55680
tcgggggaac tggcttaaag aggtgctggc cactcatccc ccaattctgt cctctacccc 55740
acagctctgg cccatggccct atagactttt ggctttgact ctctctactc ccctgtccgg 55800
tgtgtccact cccacccca tgagagttgg gagaaccacc agatgttccc ccctcccttg 55860
ggaccccaagg cactcgaa ctggagagcc tgggttcaca aggattgcag gctgcagttt 55920
aatgttgatt ctgggtatgg aaaactggag cctcagctgc acttgcgtac aaagaagcca 55980
tttggctctt ttgagggttc tcttttctc cctggatatt gaatcatggt agatatccat 56040
ttgggggatt attaggcagt gtctattagg cctactacta tgctatggcc aagaggatat 56100
gcaagaatag cattctcatc attaagaccc tccatgaccc gggcagtgcc agtgcacacc 56160
ttaatcgca gcactcagga ggcagaggca ggcggatttc tgagttcgag gccagcctgg 56220
tctacagagt gagttccagg acagccaggcc ctacacagag aaaccctgtc ttgaaaaaaa 56280
accaaaaat aaaaaatattt taaaagacc cttcatgaaa gaagagagga atagtagaca 56340
acagaacatt ctagagtgag gggctcaactg gagtcagaag gccaggtgtaa gatctaactc 56400
tgccattccct agcacaggcc ccttgggtgc ctccctctct ctgtgggtg aattggagat 56460
gccccacagag ccgtgcagtg accagtcatg tgtggtagag gattccagta gcaaaccaaa 56520
cctggatgtg aaataggaa tgaaggcgcg cagcggatca aaggaatttc aaatgtccaa 56580
tgggggtggg aagagggggg tggggggag aggcaagttgg gatcctatca agatgtcatg 56640
gggttagctgg aactcaggat tgctgaggaa cctggaaagcc agcagaggcc gaaccctca 56700
gagatccccc agcacagggtg tgaggaagcc aggtacttaa ctccagattc caccattgtg 56760
tgctaaggcc tggttctgg ggattgaaag agctgtttct gccttcctct gcccagaggc 56820

-continued

caaccatgat cctagcattt gtgtactta agcgtgtaaa gagggtttga ggaggcagca 56880
gccagatgtc tacaagagat aggcaagaca ggcttggtc ccacggcatg tccactttgt 56940
tattttccac taaccccatt ttcttccttt acttttccca ggaccaaatac agtctccctcc 57000
atgtctgagt ttgaaagttt gctcgactgc tccccgtacc ttgctggcgg ggatgcccgg 57060
aacaagaago tgcccaacgg ccctgtttt gcctttgtga gtactgagcc agtggagcc 57120
gagaaagacg ccaaggagaa ggcggggcgtt tccacccggg actgtagcca cattggtagc 57180
ttggcctgtc aggaacactgc agggagacag atgcagcgcgca gctacacggc tccagacaa 57240
acgggaatcc gagtctacta tagtccgcca gtggccggc gcctgggtgt ccctgtggtc 57300
catgacaagg agggcaagat cctcatttag ccaaggcttcc tcttcaactac cgccaagccc 57360
aaggagtcag ccgaggctga cgggctggcc gagagctcct acagccggtg gccttgcaat 57420
ttctcccgcc agcggctgga tggaggatcc gggccagca cctcgggttc cggacactgt 57480
ttccccgcct tgcattactt tgatgtcg ggcaacatga gtacgacat gaaggagatc 57540
accaactgcg tgccgcaggc catgcgtcc ggctctctgg agaggaaggt aaagaacaca 57600
tccagccaga cggtaggcgt ggccaccgtg ggcacccaga ccattcggac ggtcagtgt 57660
ggcttcaga ccgaccacc cccgcacgcgcttcc acagctggtc accccgcagc 57720
tcctcgcttg tgtctgtcg cagcaagcag atctttcctt ccctggacaa ggtccattct 57780
cgcatggcgcgccatgttgc tccggcaag tacggcgtac ccaagctcca gagacgatcg 57840
gtgttcaagc tggatagcac caaggacgc acgcctgtgaa acctgcacca gggcaagcaa 57900
aatggctccg cctgggctcg ctccaccacc acacgggata gcccctgtact gaggaaacatc 57960
aatgatgggc tttctagcct cttagtgtg gtggagcact ctgggagcac cgagtctgt 58020
tggaaactgg gcatgtctga ggcccgaacc aaacctgagc ctcccaagta tggcattgtt 58080
caggagttct tccggAACGT gtgtggccgg gcacccggcc ccactactgc agcaggcgg 58140
gaaagctgca agaaaccaga gccccttcg ccagccagct accatcaacc cgagggtgt 58200
tccaggatcc tgaacaagaa ggcggccaag gcagggtggta gcgaagaggt cagacccacc 58260
atgctgtccc aggtggggaa ggtggcattc ctgcggatg gagatggatc cttgatcctt 58320
cccagtggagg tatgggtgga cttagccctt cattcagaat gggggattag cctagaaagt 58380
tcaaaattctt ctgttttgtt ctgttctgtt ttttgtttt ctgttgcataa gtttcatgt 58440
gcccaggctttaact ttctgtgtgt gtgtgtgtgt gtgtgtgtgt 58500
gtgtgtgtgtgtgtgttat tttttttttt gtttgggtttt tttttttttt 58560
ttttttcaag acagagtttct tctgtggagc cctggctgtc cttagactca ctctgttgc 58620
caggctggcc ttgaactcag aaatcccttgccttgccttcccaagtgcc caagtgtgg 58680
gattaaaggc atgcacccaaa aatatgttctttttttttt ttagaaggaa agcaacagca 58740
gttcagggttccat agagtgtggat ttagatggtaa aaacacatgc gtacgcacagg gaagccctga 58800
attcaatccc aacacttcaa aagagataga aaggagaaag gagtgtgtta ctgagacatg 58860
tttaggtgttccat agtgggcacc atgtcacatc cttagatgc cctgtttgtat cggatcctca 58920
cgacactgag aaacacaata cattttcat tggatggaaag aggagagaaa gatttttcgt 58980
agagtgtcac ttgccttagca agtggcagag ctgagagatgttgcgttacagcaagg 59040
cttttcacca tctctttgtt cttatgataa taaaaaaatg aataatccca gcagggtggg 59100

-continued

```

ggcaaaggca ggtggatctt tgaggcaagt ctggctaca gccagggta cacagagaaa 59160
ccttgttca agaacaaca aacaccaacc cccccaccca cccactcaca aataataaag 59220
aaaaagaatg aatgtaccta aatacacatc aaaagtgtat ctggaatcct cagtactccg 59280
gatatgatag cacatgcct tagtcccagg acttgaatag ctgagactga agcaccacgg 59340
gttgaggcca gcctggacta catcgtaat tcaaaggcca gctggagcta cacagtgtga 59400
ccctgacaca aagaagcaaa cagagagtaa atttgaacta aggtgttagc tcagtggtag 59460
agcacctgcc tggcttgctc aagccctaa attccatcac tatggggtgg gggtgggggt 59520
gttggaaaat acgtatttcc attcagtaga acaccataga gcctttaaga ctaagaccca 59580
ggagctgggg agatggctca gccgttaaga gcaactggctg ctttccaga ggacctgggt 59640
tcaattctca gcacccacat gatggctcac aactgtctat agtctactt ccagggcatc 59700
tgacaccccttc ttccaaactc tacgggcaact ccatacttgc ggtgcacaaa cacgcgtca 59760
ggcaaaacttc ccaaacacaa gccaagctgg tcgtgtctgc gtagagtgtt gagaatctca 59820
acctcggaca gcgtctgcct tctctgtgcc ttgcagactt tattagaatg tggatccatt 59880
gttgccttaa taggagtaat atggtctaga gcaagatgtc agtccctgccc tcagcatgag 59940
acatggctga attctgggtgc ctgggtggct gtagcccagg cttcaggctg ctgtgtatc 60000
cttctctgtc tctccccagg atgcgtatg tgactgttagc gcccagtcac ttgcctcctg 60060
cttcatccgg ccatcccgca acaccatccg gcaactcttc tccaaagtgcg ggctgcaccc 60120
ttcagagtca ggctggggcg gggaggagag ggcagctccc cagtgagtcc ctgagcaacc 60180
aagcacccac ctcaaggcgc ccagacccctg gagatgaggc aagggtcggt gtccctcagcc 60240
tcagtcacatc caggaggaat ggcagctgtc ccactgccac agaagagctt tcacattaag 60300
gtaaagcaag gtgtcttgcg gactgctggg cagtgacctc tgattccag gggaaagaca 60359

```

<210> SEQ ID NO 11

<211> LENGTH: 76771

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11

```

cgctgagcag cgacgcccag tccggggccg gggggcccgcc ggggggtccgt acggggcagc 60
cgccccagcc cgccgcctcc ggcgcagcgc ccccgccggcc gcccgcctcc cggacgcagc 120
cgtcgggtggc cgcgctgtcg gtgggcagca gccgccttgc gctcagcgcc tcgtttgcct 180
tctccgaccc caccggaggag atgtggact gggggcccgag cggcttgggtgg cggggatgtgg 240
aggagctgcg ctggagaac gactatctca aggtggggcc gcccgggttag aaaggaggtt 300
cgcaagggtc gccggctgtc acagtccgcg gggcctgagg caccagccgg ccctccttcc 360
cactgaacaa tccggagaaa gggggccaaag ccctctgggg accactctga ggtggatctg 420
gggcctccctc attcccccccg cccctctca actttatgg ggaagcgccgc tcctgaaaaa 480
gaccttggaa ctggctacag ggccttaacc tgggagccgg gacttggaa agcttggag 540
cggtctggga ggctgtgtcg aggtgggggt gggtagccag ggaaggatgt gatctctgc 600
ttcagtcgc cagaacacctc acgcgcaggc tctcccttcc gggacgggtat tgatagacac 660
cgccgcccagg gaaacctgtg gccatccccct gcccaggaaa ggcctgaatg ccacccctgag 720
tcgttaccca gggcttcccccc gggcgccaggc ctgtctgtc ctgccttctt gggcggtcc 780

```

-continued

agtttcaaag catgacatct gggAACATCC cttggctgg agtccacttc cgcccagecct	840
aacggtccca cagcctggag gatgatgtcg tggagaacc ggTCAGAGG cagttccta	900
aatcccagga cagtgtctt tgccctccct ggggcagacc ttgtgcacac aacagccccaa	960
ttggccctc gtggccagtg ctttccatg ttcgcagtga gaggagggca actgcccacc	1020
tggAACAGGA ctccctgtg gccatgtggg gaccctaggg ctccccccgg acagcaggct	1080
ttgccttgc ttttcagtg gacatgagca cctactgggt gccagccggg ctccatctgc	1140
acgcttccag gtgccatggc agcttaggg acagcccago ttgcacgtgt tgacctctcc	1200
ctgggaggct ggaagcttcc tgagctctcc tgccgctggc tgggcacggg tttggaggag	1260
gctgaatgt tgccgactct tgcgtggta tgatggcttg gcacaacacc cagagtgcc	1320
ctggctgtgg agtcacacgc atgctgtgg tttatgtctc ccctcagcac ttaggtttta	1380
actttccctt gtggcttcc tgaccaaattt ggcttccccc tcatttaggt gttggtgtct	1440
tgctcaacta tgtctctgtc tccccatTTT cccgggata tgctctccag cctgagttct	1500
gtaaatgctt gtagtatggc aggacaaggct ctgtctgaat caagagacac cttttccaa	1560
gccaggccctc agtctcttca tccgttaat gggaggatgt cttcccaagg gcccgggttcg	1620
gtgttggccaa ctgggagttc agtcggctct ttgagtcttc agaagccggg ggttaaggaga	1680
ggcaggatgg aaggtgagga ggaggaggga ggtgacgago caggacaagg gtctttgtct	1740
cctggccctc tgggtgtctg agctgtttcc agtgagtgtc gcctgaattt gccttaaccg	1800
aaactctgag ttcttcaccc ctgectcccg gttcctgcta ctcactctgg tgaaggggga	1860
tgggttgcag agctttgttt cagggctctt aggcacccat ctccatggg ccctgtccat	1920
cccaggaaatgt catcaggctc tgacccaaag tctgttttag gCACCCATGA ctgtttaaag	1980
ggggaaacacg tcccaggagg ctgggggttg agccgggat ctcacagggg ccggggctca	2040
gctggggaca tcccacgttt gggcaagat ggaaccgtaa ttccacat tcattcagt	2100
gctcaggctt cgtgctgcta ttggatctgg ggtctgtcg ttcctgacgg cgtgatctt	2160
ggcaagttac taaaaggccc tggacctttt ctttctcatt tgtaaaaatg ggacagtaac	2220
acgctttcca acctcaacgg ttaaatgagc tgattctcg aaacaccatt agagcacctg	2280
gtatgttagta aacctctgat cactgcttt gtcagaggca cgggacatgt ttccatgtgc	2340
acccagccctg ttctctgtg gattctgtga gggcagaata gatcctcttc atcacccaa	2400
gcaggctggg ctctttctcc tgccccctca ttgggttcatt cggcagctgt ttatTTAGAA	2460
cctgeccgtgt gctaggcag atctggata cagctgtggc caaggaggaa tgaggtgtct	2520
gttttcatag agctttggg gagaccagaa ttaacaata gtcatacaag aaatcacaac	2580
acagatcata atgagggtaa agggacagt ttgcaggag cacacgaagg tggggctgg	2640
ggtgccacagg gaggctgggg gctgcaggat gcccgtggcga caactgagtg agggctgg	2700
gtaagaagat ttcaaggctaa gggaaattgg aatgctcggt gagttcaaga acagcgctaa	2760
ggccagagag gggaggagt gtgggtaaaga gaggagagca agccagccc atgcacttc	2820
tcactgtgtc catttgggtt agtcccttaa cttctctgtc tatctgtttt tcccagccgc	2880
cagcctcaca gaggcagggt caaaatccag agactaaatg tggaccataa atagaggggt	2940
tattactggc ggtgttgtgt tggatattcc acagatgggaa aactgtggc cagagtgggt	3000
cccagagcag ggtcgctagg ctgccaggc ttggagtaag catgtgttag actcttgatg	3060

-continued

tgttttctgc	ctttctggac	tactttgct	ctaagatgca	gaaactgagt	acaaggaaa	3120
ggactccaaa	tccccagtagt	gtccccctgca	ggagcagagt	tgaaggaggt	cccaccagg	3180
ctgggtgggt	gccctgctca	ggcttacctc	ccccttggtag	gggtcactgg	gaactgtcaa	3240
ctgcgaagga	ctgttatctc	cagtctggta	tggtctgctc	tggggtccaa	gtgagggaaag	3300
cataaatcg	gcaactctct	cccagtgaga	agatagaagg	agtttctaca	gagaaggtag	3360
tgagctca	gtcactttaa	ttattcagag	agaggcggga	tgaccttgca	tgcttggata	3420
aagatagagg	gtttgtggaa	tgattgctt	agaagggttc	agagggagat	tgatgtgg	3480
gctgcagtt	agctgcctga	ccttgggtcc	cttctagcca	aggaagtct	aaggttatta	3540
actactgagg	tgccaggatt	ctaagatgca	gcttagctg	ccgagcacac	acgtcacctg	3600
agaccttcaa	agacccggtg	atctgaggtc	ttccagccct	tcatcaccac	taaccattaa	3660
gcacgtacca	tgtggccaga	acatgttagga	accactggg	tgtggatgga	ggaatctgag	3720
agttggagtc	tgaagacctg	actttgaaac	ctgactctcc	cgttactgt	gtggccatgg	3780
gcaagttacc	tggcttctct	gagctgctt	ttttctcatc	tgtataataa	aggtgatgg	3840
gcctggcgca	gtgactcacg	cctgtatcc	cagcacttgc	ggaggccaag	gcggggcgat	3900
cacctaaggt	caggagttcg	aaaccagtct	ggccaaacgtg	gcaaaacccc	atctctacta	3960
aaaatacaaa	aattacccag	gcgtggtcgt	gggcgcctgt	aatcccagct	actcaggagg	4020
ctgaggcagg	agaatca	gaccctggga	ggcagaagtt	gcagtgtatcc	gagatggcgc	4080
cactgcactc	tagcctggc	aacagagtga	aactcttgc	tcaaaaaaaaa	ataagtaaag	4140
gtgatggaac	ctgcctggac	aacctcatac	gatggtcatac	aggaatcagg	gaagtgtgaa	4200
tagacaagaa	aacatttgc	aagcagtgaa	cccaagtcgc	ctgaggaatt	gttattattc	4260
atagccaata	ggcagacagg	atgtctgtgt	cttcattctt	ggaggaactg	ggagaatgca	4320
ggagggtgga	ttcccttagt	ttttagaatgt	taatacatca	gtagctgcc	cagggagaac	4380
ggaaagggtgt	gatcaccttc	cctgcccgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	4440
gtgtgtgtgt	gtgtgtgtgt	gtgtgttttc	cctccttgg	gattctagat	gtgagtgaag	4500
tgcagccagt	gtgtgtgtgt	gtgtgtgtgt	gtgtgttttc	cctccttgg	gattctagat	4560
gtgagtgaag	tgcagccagt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgttt	4620
tccctccttg	gtgattctag	atgtgagtga	agtgcagcca	gtctagatgt	cagctcagaa	4680
gggcctggag	gaagcaccta	atccatcccc	acttgcatt	ttacagatca	gtaaatggag	4740
gctctaaatt	acatagccag	gattccaacc	cagggttgc	taactctggg	cctgttcttt	4800
ctgcagtgt	accagagcag	tgagcttttgc	gttgttaact	agcgtgtct	ggctgagaca	4860
catttggta	gaggaatcct	gtcatggcc	aggctctagg	ctaaggcatg	ctgggggccc	4920
ggagcctggg	ggaggtgggg	gttagtggag	gtatgttgc	tgtgcattgt	tgtgagatgg	4980
ttccttctgt	aacatggggg	acaagcacgc	ccgcccagg	ccgtgaggac	cctttcaat	5040
gtgaggacgc	agatggagtg	cacagtgtgt	gttagtgcgt	cccagtatt	gccgatccct	5100
tatccgttcc	tccaggcatt	cattcattca	ttcaacacat	attccccag	tggetcctgg	5160
gtaccaggcc	ctgtgtgggg	tgtctggacc	cagtggccac	gttagtgcgt	tctgcattgt	5220
gctcttctct	tcttggctcc	caggagctac	agggggagga	ctctagtc	tcactgggt	5280
gaggagagga	ggtcaccaca	gtcagcaaag	ctcttggct	ctccttaccc	tgagtcccag	5340

-continued

ccgcagcctc	cttggggctg	agcctgtgtc	cctgggcata	ctctgttgc	cacatacaaa	5400
tgcacatata	ctgttcacag	ctcaagctca	gagacacaca	gccacttgca	cacatctcac	5460
agacccatcc	agtgactcgc	tggtgttgg	tgctttcaga	agcacgcctc	cagacagatt	5520
agagaggata	ctcaactcaga	gaagtgcaca	aaccatgca	gccacacaac	tcccattcta	5580
tcacacgcac	acatcaccaa	acagacacaa	cgagggtca	gaccctgctg	caaccccccac	5640
gtcagaaaca	cattctcata	tacacatctt	acaccctag	acacacccgg	atatgtcccc	5700
gagacagact	gacatcctgc	agacacactc	agacttatca	gaagtacaca	tgttagggcg	5760
gacacaaacc	cacacagacg	ttatacacac	acactagcat	ccacagaagt	ctgcacatct	5820
acccttgcca	ggctggccca	ggcacaagct	caccacccgg	gtacctcccc	tgcctttct	5880
acaaagagag	cagcagagac	ttgcacacct	acccgaggct	tacagacacc	caccaagggg	5940
actccacat	gcacatcagc	agggacacag	cacagtctgg	catgcaaaca	ttaggcatgt	6000
gcatttgcc	cacaggagcc	agaggcttc	aaaccaggga	acctagacac	acgcacacac	6060
gtgcgcac	gcacacacac	acacacacac	gctgtgacat	ttcatccaca	cacagaaaca	6120
ggaaatcctg	gtctgtgtct	ccctggagct	aatgtgtaca	aactactact	gttttcagtg	6180
gctcaactaga	cttacttcta	caaggctgct	gggaacacta	aggagtgtgg	ccagggcctt	6240
gactttgtaa	cttctggag	ctgggggtct	tggggtcctg	ggtgggtgggt	ggatgggggt	6300
ggtaggtctt	gaagtcagaa	gaagtcacac	cgaagtcccc	caggtccctg	atgeccagct	6360
cctggggccca	gggccagtg	ggctgagtga	agtggaccga	ttgtcgagct	gcccacagaa	6420
gaaacactgc	ccggggggca	ctcccttcta	atctccttg	aatttcctgc	tggccctatc	6480
tcccccttct	ccccctgttc	ccttgctgg	ggctgcagag	ggccatgcca	ctcccatctta	6540
taaagggtgc	cccattctgg	ctggggatg	atggaggcaa	gggttagaggt	ggctagagga	6600
gctatgggag	ctcagaggag	ggacttggca	gacttcctgg	aggaggaggt	gtccttggag	6660
ctggggagtt	agoctcaagt	ggctgtctct	gtgcccagct	cttagaacag	tgcctgccc	6720
gtttaggggg	ctcaataaaat	atgtgtacaa	ttaagtgaaa	ggcagtagggc	agggcctgg	6780
ggtggaccac	aggcttgaaa	cggaaaagct	ggagggagtg	gtgggagctg	gcagggatgg	6840
ggagagtctg	ggaagggcg	ttgggctggc	ttgggaaggg	cattgaatgc	caagccaagg	6900
gagcacctt	ctggggagtgg	tcttggca	gcaaagctgc	agccacatcc	ttttatgtgt	6960
ctgcagccct	gggcctgagg	cctgcctgc	cgagaggcc	tggggaggg	ggagagccag	7020
gctgtgtgg	ctagggggca	actccaccc	tggagtcaag	tctggagcca	aacctcaactg	7080
tgtcacctcc	tccgtgtca	tgctgtccct	gtgcaaacc	ctcaccaaga	ctccgttttc	7140
ctccctgtaaa	atggggatgt	gateccttct	agtggggcac	ttggggaggct	gacagctcag	7200
atgttctgt	cacctagcac	catecctggca	cagaggagg	gaggaggccc	cgtgtcttc	7260
atagcctccg	gggcagagag	agtcccctag	gaaagggcg	cagctaagtt	caagcaaggg	7320
cctgcccage	agccaagggtt	ttcaactgagg	tcatccccct	agtccccaca	tgcacatgc	7380
catccatgg	cacgttact	ccccttgg	ccggaaggaa	cttggagctg	gggtgtctca	7440
tgccaggctg	accctggcat	cactggaggc	ctcagtcagc	actgcacatca	cctgctttgc	7500
ggtagtgagg	cacagcaggg	gtcagagcag	gttggaaagcag	caggccagcc	ctttgttctc	7560
cgaggtccag	ggtggttggg	agagagcact	gtgatctgga	gccaagctga	ctgggctgg	7620

-continued

atcccagctc	tgcctttcc	tttctgagct	ttcccttct	gaggctcggt	ctcgcatct	7680
gtagaatggg	gatgtggatg	tccttgcccc	gcacgttgcc	gtgagggttc	cgtaaagagaa	7740
agtgtgcagg	tgtgggtccc	attagccagg	tctggctct	ggaaggacc	cgacagcact	7800
tggcgcttgc	ctgttatatac	tggctctgg	gtgacccccc	tgacccagg	ttaagctgg	7860
tgcttgatga	caggggctgt	gcttggctct	gggtgtccct	ctgtgggtgt	tggcttgggg	7920
catcagtctg	agtgttatgc	acagagctgc	actgcattt	gggcacccccc	aaacgtata	7980
ccctgtctgt	gtccccccctt	ccctccccca	cgcgctcg	tcactcttcc	ctgaatggcc	8040
caggttagt	cctccagctt	agtaactggt	gatacacatt	tgctaattg	aggacacett	8100
tggggcccc	tcattctcaa	atggaggtga	atcatctc	actgcctgag	ccacaaggaa	8160
aagctcagcc	tctatgaaaa	aaattctggg	gttccctago	taggctggat	cagctgaaca	8220
tctttactct	ggttggggcc	caggcctgct	ttgggtcttc	tgggctcgag	tcagttgcac	8280
cagcctctcc	cctcccccgc	tctcttgc	cctacccat	ggtggagcag	cctgtgtgg	8340
cctgectccc	acagcctcac	ccacccctgc	cccgcccaag	ctctgcgggt	gctggacaca	8400
ccccggccctc	tgtccctgtc	tcttgcttcc	cccatccgg	cttgcgtctt	gcagttcccc	8460
gctccttccc	ttgggttctt	cctccaccc	tgagcgctca	gccccagcag	cagctcccc	8520
ggtcaggctc	ctgtgtcaga	cacccctggag	ctccctggag	cttccttgg	ggcatttacc	8580
acgatgtga	tttagattatt	aatggggtaa	tgataggttt	catgcatca	ctgccccacc	8640
tttccctgca	tgagggcagg	gacccctcag	ggttaatttgc	tccttcctt	tacaaaggga	8700
gaaactaagg	cccagagagg	ttaaggggct	tgttagaggt	catgcagcct	gtaagtggca	8760
agaggcaaga	ctggggtctt	gtgtttctct	ccaggatgcc	aggcaggggg	aatggcaggg	8820
gtgtgaagt	ctgggggtgtc	ctggagcttc	aggcttgcgt	gccttggcc	actctcttgc	8880
gagcttgc	ttctatccct	aaagacctca	gaccctgct	ctgagttttt	gagatttgc	8940
tgcaacccag	agggtgtttt	ggggatgttt	taaacacato	caggtccaga	tgctgccttc	9000
cttcctgttc	acatggaaag	gatgggagga	aactgtcagg	cgcttctg	cacgaatgcc	9060
ccagaggccg	tagggcctgt	ggatatctcc	ctgctctgc	aggtaaggcc	ttgaatggg	9120
gagtgacaac	agtgacaagc	cagatgtggc	cttgggtggc	agtggctct	taccaacaac	9180
aaccaccaca	cacacacaca	cacacacaca	cacacacaca	cacacacccc	tccctgata	9240
agcatctgta	gtgtcttttgc	tgcagaactg	taccagacgg	atctggagg	ctcactcctg	9300
actggctgcg	tagacccat	cttcctgggc	ctccgttgc	ttgtctgt	aatggggata	9360
aaataataat	agtgcctgcc	tctaaggatc	atcaactg	aagagataat	ggctttctg	9420
tctcagtccg	gtgtctggaa	cgcattaaac	gcttgcata	gttagctatt	gttatcgta	9480
ccattgttca	ttataaagact	gtcttagtca	gctttagg	gctccaaactg	gacctgaatc	9540
ctggctgtgc	cacttagtgt	gtgaccctgg	gcaacatact	tgacccctgt	gaggttcc	9600
tcccatcggt	gattgtcg	gctgcgtga	gagaatgtgc	actccctggc	ggggatcagg	9660
ggtgacacagg	gagcatctca	tttcttggca	ttggagaggc	gggtccaggc	cccagtctt	9720
cctcattatt	cccaaccaga	ggagcaggac	gtgggtgatgc	tccaggccca	ggcagccct	9780
gtgttgtctg	gagtggggac	atcatgtggc	acgggtatgc	caggctgcgt	gtctggcc	9840
tctttccccc	ggacctttttt	ttttttttt	ttggagacga	agtctcactc	tgtcgcccaa	9900

-continued

```

gctggagtgc agtggcgtga tcttggctca ctgcaacctc cgtctctgg gttcaagtga  9960
ttctctgcc tcagcctccc gagtagctgg gactacaggc atgtgccatc atgcctggct 10020
aattttgtta ttttttagtag agacggggtt tcatacatgtt ggccaggctg gtctcgaaact 10080
cctgaccta ggtgatccac ccgcctcaga ctcccaaagt gctgggatta cagggtgtag 10140
ccacatgcc cgcccttcata aggacaccttgc atgcaacttc ctgaaaactg gacccttagct 10200
tcttcgcaag gatgctaagc tagattgagc tctgggggtgg gagacaggtg cccagagtcc 10260
ccccccac tcagtctat cccctccagg gactgtggga ttgtccctgt tccttctcaa 10320
ggcctttctg tccctgggta tgaagggatt gagttttggg tttccctccc ctgtgtctgt 10380
gcccagagtgt tttgcatttg ggctgagaag tggagctctg cggagtttt tgcatatcct 10440
ttttggggtt ccaagtgtga tcatctcata tggagcccttgc acagactctt gggggaccgg 10500
cagaagggac ggcagatggc tgccctggcc tcttgggtgg ggctggcatg ccgcctctg 10560
ggccactgcc ggcccaactcc acagagcagc cttccgggg ctggggtag tggcccgccc 10620
ccgggcagga gacgagtgcc ccagagccag ggtggcagat gggaggcagc ccaggccctcg 10680
gcctgaggga gggaccatgg tggccaggcc agtatgggag caggcggggcc aggcggaggc 10740
tcccgttccct ctgttagtgtg gttgtctgtc cgcagagtga aggttctggg cctctcagct 10800
cagctctgtc attcttctgg gttgtctgc ccgtcgagg gcttggggg cccctggAAC 10860
ctggggagga ggtgttacat ttgggtttagg tttggatct ggcagggtctg tggccctgggt 10920
gttgttgggc tgggacttga gccccaccac tgggttctga agcacccatg ggcatttagct 10980
tcacagaatc aaggaattttag attcaaacaac aggagcgtctg actagaatgtt ctgtta 11040
ggggagtaga acagggggat tcttccccc tctgaggggcc tttgttattca ttcagcacct 11100
ctttattttagt ctgttctgtg ttttttttttactgtgttgg gctctgttctt cctgacatca 11160
catcctgggt gtagagactg acattctgc gggaaacaga caaaataaaac gctgagtatc 11220
acaaataaagc ctatgatata tgccaagaaa gaaacaaggg attaagatgg acgtaatggg 11280
cctggaggag catggaaaggc ttctcagagg aggtggcatt tcagtgagaa cttggaggcag 11340
aaggcaagcc agcaggccag agatgggggg cagggaaatc caggcagagg aaactgcaag 11400
gccaaggct ttgagggtggg acagggttggg tggctgttgg ttgcctacga cccaaatggc 11460
tgctgttggc ccagttagag gagggttgc gggagatgaa gatggagaag tggacggggc 11520
tttaggcagg cagggttgcgt atgtgttgc atggatggg ggtttttttt tttttttttt 11580
gaaaagtccac tggagggtttt gttttgtttt gttttttttt tttttttttt tttttttttt 11640
tgtactgggt atatatatac aatataactgt acaatataact gtacaactat atacaatata 11700
ctgtataata tattgtacaa atatatacaa tataactgtac aatatattat acaaatacat 11760
acaataaaat gaattttggc aaatgcaccc agttctgttag ccaccaccac aaccaagata 11820
tagaacagtt cttcacacc ccaaatttca ccctgttggg tcaatccctc ttcgttcc 11880
agaccctgac aaccaccggg actgtttctt gttcccttagag ctttggccgtt tccagagtgt 11940
caaaaacaga ctcatacagt gtgttagccaa ctcagttttttt gttttttttt tttttttttt 12000
gcatgttggc ttccctccatg ttgcagggtgt cagcagttctt tttttttttt tttttttttt 12060
gtattctattt gctgtggacgtt accacagttt atcattcacc agctgaggga cttttttttt 12120
atccccaggt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 12180

```

-continued

aaccacattt taaacatgg ttttagaaaa atcactca actgtaatg gagaacaat 12240
tagttggag cagagtgacc atcataagga ttaaatccaa taatttggg atggatctgg 12300
cccaccatag gcactcacta attgttcaaa caaaagggt gtgcctggg acagtcagg 12360
aggctttgg tggaggtggc tttttttt cttttttt tggtaaaca cagtctcatt 12420
ctgtcatccc ggctggagta tagtggtgc atcttgctc actgcaacct ccacccctg 12480
ggtcaagca attctccgcc tttagctct gagcagctgg gattacaggc gtgcaccacc 12540
acgccccgt aatttttgtt attttatag agacgggtt gccatgtta ccaggctgg 12600
cttgaagtc tgacctcaag taatccgtcc gtctcggct cccaaagtgc tgggattaca 12660
ggtgtgagcc actgcgcctc gctggaggtg gcttttgagg tgggaaacag gtgtggagta 12720
ggagtcctcag cgactacaga gtccatggg agtgggtggg ctcagattcc agctcccg 12780
ccctgagcca gccttggcat atgttagagtt gctagaatat ggtctgttagg ggtcatatca 12840
gagttccca gggacactag gggctgaatc ttgggtggg ggttagcttag ggacttccg 12900
tctggaaagtc ttttccttgg tggctggcaa gaggcaacct gctcctgcct ttggagect 12960
catgtccccct tggttctctc cgtgtccctcc cagctggag ctgtgtgg gttcagacca 13020
catgctgcca ctgtgtttc tgcaagggtt cctgggtttt ggtcaacttc atctccagcc 13080
ccacagctca ccagctgctg caggccattt ccgtcatagc catgcacatcg gagccacgt 13140
caaagggtcc agtggggccc aaatctaaac ttccccccgc tgcctcagt gtacaaacgt 13200
aaatcatttt atatctgcgt gggttctga tagaaagtaa tctttcttag aacacttaaa 13260
gtgtgattta ccagccagg ctggtaaattt actaaagcaa gacccctct ctacaaaaaa 13320
tacaacattt agccaggcat ggtgggtgtt ggcagtagtc ccagctactc aggaggctg 13380
agtgggagga tgacttttagt ccaagaggctc aaggctacag tgagctgtga tcaagccact 13440
gtgctccagc ctgggtgaca cagaggacc ctgtttaaa gaataaaata agtaaaaagt 13500
gtgatttcaa agtggagctc tgcttctccc gctgttattt cattcaaca tttaaagtt 13560
tatcttattt tggctgggtg cggggctca cacctgtat cccagcactt tgggaggccg 13620
aggccgggtgg attacctgtc aggagttcaa gaccagactg tctaaccagg cgaaaccccg 13680
tctctactaa aaatacaaaaa attagccaaat cgtgggtctg catgcctgtat atccagacta 13740
ctcgggaggc tgagggtgggaa gaatcgttt aacccctcag gcaagggttg cagtgagctg 13800
agattgcacc attgcactcc agcctgggtt acaagagtga aactccgtca caaaaaaaaa 13860
aaaaaaaaaaag gaaaccaact atattgaaac acagtattttt aagtattttt aaattacaac 13920
atagtaatat gtgtgtttt taactcataa attctggta cagatctaact tagcataatt 13980
tcaaaactact gatgagtgta aataatattt tgacattttt gcagcaactg ttatatgata 14040
tgaaaatacc tggatttttt ttttgggtggc aaagtccacag atattactca tataatgtg 14100
atttgttcat aggcccagaa tagaaggaaa tgctacattt tacttagagg tcagtgaaaa 14160
ggaagatgtt attttcaag ttcatggact ccctgagttc tggttcaag gaaagatctt 14220
gaagttgggtg gttggacttt gagtcctgtc tggaggtctc accccctccc cctgcctgtt 14280
tgcgtgtct tggccaccac ccacccccc tgcgttctc tgaagttctt catctgttaga 14340
acatggatgtt tggttctgtt ctttccttac tgggaggtggg gggttcagtg aggatctgt 14400
agatgattca gatgaaatgtt ctttgcataac tgtgaagccc tctgctgtc tgcaagatct 14460

-continued

```

tgctattatt agcagtagtg tgagctatg tttggctgtc cggcccgacc gggaccaaga 14520
ggcttccct ggacaattct cccaggattc ttggaaatca gtccatcgag gggagcagg 14580
ttgccttga gaggctgctt cctgcaggct gagctcatgg tttggcaggg ggtggacaga 14640
gatgacacctt gtgggcaggg agtgggtggg tgagcaggag gttccctacg tagacagacc 14700
ctgcctaccc tccttgggtg gcataggat ctgcgaccag ttctgatctg cccaggctct 14760
ccctctctca tttttatcta gagacagggg tctcactctg tcacctagcc tggagtgcag 14820
tggtgactc atggctcaca gcagcctgga actcctggc tcaagtgaac ctccctgcctc 14880
agcctcctga gtatggggg actacaggca catgcccagc taagtttta aaattttta 14940
ttttttgtta gagacggaga ctcaactgtgt tgccaaggct ggtcttgaac tcctgtcttc 15000
ctaactcttt cgagctggct gggagttgtg gactggcagc gacagaagct atactctttt 15060
ttttttttt tttttttagt cagactgtct ctgtgtcaact caggctggag ttagtgggtg 15120
caatccctgc ttattgcgc ctcaacctcc caagctcaag tgatccctt gttgcagecct 15180
cctgagttatc tgaggctaca ggtgtacagc accatgcctg gctaattttt aattttttt 15240
tagagatggg gtctcgctat gttgtccagg ctggctcaa actcctgagc tcaaggatc 15300
ttccctcttt ggccctctgca aagtgttgcg attacaggcg tgagccaccg cgccctggccc 15360
cataactcatt cttaattcc tacatctacc catagacagt gacaggtacc tgctctgtgc 15420
cacggctctgc cctgggtgat gctgaggact agacacaaac cagccctgc ccttccct 15480
agctcttggg gaagattgac atgaaaacag acaaatacaa cctaaaatta ccagcacaac 15540
ggaccataga tttcagagtg tggggtttga atcagagtct gggaaaggaaa tggatggaa 15600
ggatgagata gtcaggagtc agcctggcaa aggccgttct agcagaggaa atagctgaac 15660
aaatttggaaag agcacagttt tgacgcactg acccaggggcc acgtggcttg tttggatcc 15720
agccccgcag gtcccttgct caagctgtc tgcaagctt ctgagcctcg tctccatc 15780
tgcaaaacag gaacgttttgc agcacctacc tcacagggtc ggtaagagga tttaactgtc 15840
aactgataat gcacgtactg cctgtgcggg ggcctggtaa ttatgtggcc tcctaatagg 15900
tggtagccgt gttttgttg ttgaaacagt ctgtctctgt cgccaggct ggagtgcatt 15960
gttgcataatct cagctactg caacccctgc ctctgggtt caagcgattc tcctgcctca 16020
gcctcccgag tagctggat tacaggcaca cgccaccaag cctggctaat tttttgtat 16080
ttttttgttag agatggggtt tcaccatgtt ggtcaggcaa gtctcaaact cctgacccatca 16140
ggtaatccat ccgcctcgcc ctcccaaagt gctgggatta caggcgtgag ccgcctgtgcc 16200
tggccaatgg tagccatttc taaagtcaaa gcttatttg acgctggagg gctgggttaa 16260
ggcgggagta gggaggagac caggggctgg ctcaggaccc ggtaaaggaa ctggaccta 16320
ctgcagaagc tggtgagaac cggtggaaaga gatagaaagc cgtctacctt cttagattt 16380
tctccaaaaa cagctattaa ctgtaatcag agcaaaaaag cattctctag aaagggaaac 16440
actgcctatac ctggcagcagc tgggggtgcc ctggcccccc atgccttat 16500
tccccgtttc tagtccagct aaaaaatctg ggttagctt aagagtagcc tttaggtg 16560
agaggtcaac ttgaaggagg tggccaggcc gttggggggc tgagggtcag gggcccaag 16620
gtctgaggt gggcactacc aggcactgtg ccctttgtt gtctggcaga ggtgaggct 16680
tggggtttcc ccagtgggtgt tgaggggatt tttttttttt ttttagaca cagtctcgct 16740

```

-continued

ctgttgccca ggccggagtg cagtggcg atcttggctc actacaacct ctgcctccca 16800
aattcaagca attctcttgc ctcagcctcc caaacagctg ggattacagt cgtgagccac 16860
cacagctggc taatttttat attttttagta gagacaagggt ttccacatgt ctacaacttt 16920
gttccatagc cacgaggcct actgagtcua gtggagctgg gcctctccct ccccacttcc 16980
acacagttag ggggtctggg agctggaga ggcttccatgc aaccactgag ctgtctgtga 17040
aggagctggg aagcctgtgc tcctcaccctt cacctccag ggccccccac cccagctcc 17100
ttcagttgga tctgtcaactg gagagcttc ctgggatggaa ggcccaact gcatgaccac 17160
agagccccag aacatttta ggctccctga ggactgtgaa cctctggaaac agaacaaagg 17220
cctgcccctt gggtcccttc cccaagcccc agcagcacct ctcttagccc cacctcggc 17280
ccggaaagcc cagcgccccctt gcctaagtcc tgcatcgcc agcctctccc gcccacgtcc 17340
ctgcagaccc agoctggggc cacaaggctg ctttgcacatg ctaccctctt cacgtggggc 17400
cgtgccagca ggggctaaacg ggcattctgc ctgcacgtgg gactagcccc ctgagaggcc 17460
tagggccccc tggagagatg gaactggcgt gaatgcaaca tcttgggctg tcacaaaattg 17520
actgtggta agggcaggcc ttggggagac agctgctcat gtgagaacta tcattcattc 17580
attcaatttat tcaaaggatg cttgctgaga tcctgctctg tgcaggcccc tggcatgggg 17640
acttgcgtctc agagacctga ctcttgcctt tttagggggg gaaataacta gaaggaaacg 17700
cccttcatca gaatccctt gggggctac ccattggatgt tctgatttcg caagtctggg 17760
ttgggacatg agaatttctc acaaatttac cagtgtatgt gatgttactg gtccaggaac 17820
tgcactttga gagccactgt aattcaggtt ttgtattat gcatggatta tgtttcttt 17880
ttcttttctt cttcttctt ttctttttt ttttttttgg agacagaatc ttgtctgtc 17940
acccagggtg gagtgacgtg gtgtgatctc agctcaactgc aatgtccatc tcctgggttc 18000
aagcaattctt cctgccttag tagtggat tacaggcacc tgccaccatg cctggctaat 18060
tttatatttt tagtaaagac agggtttac caagttggcc aggctggctc cgaactccctg 18120
acctaataatg gtccggccgc ctcggctctcc ccatagtgtc gggattacag gctgtgacca 18180
ctgcacctgg cctgtttctt ctatttcaa aactgtctctt atgtatgtt tttactttta 18240
gagacagggtt cttgctctgt tacccaggctt ggagtgcagt ggtgcaatca tagttcaaca 18300
gaaccccaac ctcctggctt caaacaatcc tcccacccca gcctcccaag tagtggggc 18360
cacaggtgca ctaccagacc aggtaatcca tccacactcg cctccaaag tgctgggatt 18420
acaggcatga gccacccgcac ctggccaatg atagccattt ttaaagtccac ggcctatttt 18480
gatgtataac attattatta ttgttggat tattattattt attattattt tttttttgtt 18540
gagatgaggtt cttactgtgt tgccaggctt ggtcttcaac tcctgggctc aagtgatect 18600
cctgcctcag cctctcagag tgctgggtt acagggtgaa gccactgtgc cctggccagtg 18660
ctttatTTTA tttttatTTTA tttttttttt ttctgagaca gagtctcgct ctgttaccca 18720
ggctggagtg cagtggcatg atctcagctc actgcaacctt cccctctgt ggttcaagca 18780
attctcctgc ctcagcctcc tgaatagctg ggattacagg cggtgcaccac cacgccccggc 18840
taattcttat attttttagta gagatggggt ttccacatgt tggccaggctt ggttcaaac 18900
tcctgacatc gtgatctgcc ggccttggcc tcccaaagtg ctgggattac aggctgtgagc 18960
catcgccccc gacctggccctt gtttataaaa taagaaaatc agtcatttaa aagttaaaaaa 19020

-continued

aatgatttc caaaaatgac tcttagtgctc tgggtgtat cttgtgtaa agatgttacc 19080
tcccatgatc taaatgtcat gcattgtta atgtagcccc gcatcatagg gttgtgtca 19140
tttcccccttg gttcaacctg agcttgggtt ccacttagac ctccttattc ttccatcac 19200
catatcctgt catttaact ttcttagagat aataagttag gtgctggtaa ccggcatggc 19260
taggatttga acacagatcc atctggaggc tggcttagcc taccgggtt cacttcctga 19320
gtgatcttgc ttccatgctt ctttcacccgg ctgcctgaaa cagcgcgtgg atccttcact 19380
tcctcatttt tttaactggt gtgatcattt actgttccac ttcttcctct ccactctgg 19440
tctgtccacg tggtatgctg gtgggggtggt gagtctatga gggtggaggt ggggggtgggg 19500
ggctggcaaa gtgactgttc tctgcagccc caggagtcaa accgatgggc aggaaaacct 19560
ctcgagcttg tgagaactgg cctggaggcc agggctccct aaccctgact ctgagttaaa 19620
tccatccago ctcagggggc gctggcaacc ccagtcgtac ccccttcctc acagggagcc 19680
atcctaccag gctgaccatt ggaagagcag agcagtcctc cccatctgg caggcaggag 19740
tgaagagaga gaaccccatg tttgttcctt tcacaggcac aaaatcagat ccccttcct 19800
cccaaaaaa gcgagggtgt ggtgggggt gctcggacag aagccagagg tgccctctg 19860
acggggcag agtggacact ccctccccc agtggcagca tgaggaggaa ggaatacgg 19920
accccaacttc ccagtcctt ccttcgggtt ttctttgggg aaagacgggt agggagagct 19980
gaggaatccg gctctccaga tctgcctct ccaggccca cagagccctt cttgcctt 20040
ggaccccca cccctgttc catcaactgcc atcctcggtg ctccccctgg ctccctcagg 20100
atgcctttag caggtcatc atccctgggt actctgttcc ttctttttt tattttgtga 20160
gacagaggct cgctctgctg cccaggctgg agtgcagtgg tatgatctt ggtcactgca 20220
acctctgcct cccaggttca aatgatttcc ctgcttcagc cttccgagta gctgggatta 20280
cagggtggccca ccacatgcc tggctaattt ttgtatttt agtagagacg ggatttcacc 20340
atgttggcca ggctggctt gaactcctgg ctcaggcga tctgaccgcc ttggccccc 20400
aaagegctgg gattacaggc gtgagccacc gcacctggcc agttccctt cttagtatac 20460
aatcttattt aaatcattcg gtcacattaa gagtaggtat tatcattccc atttggact 20520
aggtttacat atttggatga gacataataa ttatgcattt ctctgcgtt attctacttc 20580
cctcccccctt acattttagta tctaatgatt gttgcacatt cgtatgttctt ttttgggtt 20640
tttgacctga ctaaacatgc atgcatccat tcagttata ttgttgggg ggctaccacc 20700
tgcttagggca aatgggggtt ctcaggcga ctcatggttt ggctgggaga caaactattt 20760
aacatagaat tgcaaattt ctaataatga ttacaaggaa gaagtggagga acagctggaa 20820
gctgatgaca gcggtacgaa aagctgcattt caggaagtga ggtcataactc ttacccttc 20880
tgcttaaaac ccaaaaccag gccaggcatg gtggctttagt cctgtatcc cagcacttgc 20940
ggatgctgag ctggcagat tgcttgatc cagttcatga gcaacctggg caacatggca 21000
aaaccctgtt tacacacaca cacacacaca cacacacaca cacacacaca cacacccac 21060
cccccaaaaa ccatgtgggg ttgtgtgtt ttgggtccaa ctacttgggg ggctaaagggt 21120
ggagggtggc ttgagcctgg gaggcagagg ttgcagcggc ctgagatttt gccaccgcac 21180
tccagcctgg gtgacagagt gagacccat ctcacaaacaa aagcaagcaa acaaatccca 21240
aacccaaattt ctatggcc ctccacagta ccatgtgggtt gactcctgtc agtttcctca 21300

-continued

gcctcaaatt aacccagtag ctgccctggc ctccttgctg taccttgaca cactgagtc 21360
tttccctgcct ctgagccctg gcacacactg ttcctctgc ctggaataac ctctccccct 21420
agctttctcg gctgttccctt cttgtctcag ctcaaatgtc tctttttag agatggcctc 21480
cctgatcatg tcccctaaca tagcacccccc cctcacccta tcatataact catgttgttt 21540
ggtttcattt tggctttgtc tttatagcac ttaacagtat ctgagggttg tttttgtttt 21600
tttttcttgg agatagggtc tcactctgtc acccaatggt gcaatctcg ctcactgcat 21660
cctctgcctc ctaggttcaa gtgattctca tgcctcgcc tcccgagtag ctgggattac 21720
aggcgtgaac caccacgctg ggctcatttt tttttttttt tttttttttt ttttgttattt 21780
ttagtagaga cagggtttca ccatgttggc caggctggc tcaaactcct gacctcaaatt 21840
gatccaccca cctcgccctc ccaaagtgc gggatttagag gtgcgagcca ctgcgccccag 21900
ccaggtgtct tttgttattt ttatcttctg gtttactgtg tatcttctg gtgaagcttg 21960
gggctccata cagagtatat agaaggggct ggacccaccc cattgtccct ccctgcccc 22020
gcgcttagca cagggctac ctaggacgta catagttgtat gctcatccat gtttctttt 22080
ttttgagacg gagtctcgct ctgtcgccca ggctgggtg cagtggtgcg atctcggttc 22140
actgcaagct ccgcctcctg ggttcacgccc attctcctgc ctcagccctc cgagtagctg 22200
ggactacagg cacccgcccc acatgcccagc taatttttt tatttttaat agagacggat 22260
tttcaccatg tttagccagga tggcttctgat ccaacactcgat gatccgccccca tctcagccct 22320
ccaaagttct gggattacag gcatgagccca ccgtgccccaa cctcatccat gtttctttaaa 22380
tggctctcc tggttccccca gtgtgttctg gttgttctg taacaaatca ctacaaatttt 22440
agtggcttaa aagaacacca ggccaagggc acgggtggctc atgcctgtaa tcccagcaat 22500
ttgggaggcc gaggcaggct gattacgagg tcaggagatg gagaccatcc tggctaacac 22560
ggtaaaccgc cgtctctact aaaaataaaaa aaaaaaaaaatt agccaggcgt ggtggcgggt 22620
gcctgttagtc ccagctactc aggaggctga ggccggagaa tggcgtgaac ccaggagggg 22680
gagtttgcag tgagccgaga tagegccact gcactccagc ctggcgacca gagegagact 22740
ccgtctcaaa aaaaaaaaaaa aaacaccagg ctggcgaggg tggctcatac ctgttcccc 22800
cactttggga ggccaaggcg ggcagatcgc ttgaggtcag gagttcaaga ccaacctggc 22860
caacatggtg aaaccccgtc tccactaaaa atataaaaaa tcaagctgggt gtgggtgg 22920
gcacctgtaa tcccaactac ttgggaggct aaggcaggga gaatcgcttg aagttagggag 22980
gaggaggttgc cagttagctg agattgtgcc actacactcc agcctggcaca acagagttag 23040
actccgtctc aaaaacaaca aaaaatgaa caccagatta ttatcttctg atcctggagg 23100
tcaaaaggctc agaaaatcaag gtttggcag ggctgcattc ctcggaaagg ctccagggg 23160
gaaacctttt ctttgccttg agaggcaccc ctttccctcg tcatcacagc cccttcctcc 23220
aacatcaaag catctactc cagccctctgc ttctgtttcc ctgttgggg atccttaatt 23280
tactcatatt tgcaaaagtct cttttgcctt gtaaggcaac attcacagat cctgggaatg 23340
tgaacatggc cgtctttggc ggtctgtcat tcaagectgcc acatccaggt cccccacgg 23400
gcttcattca gtccttaggtg tggccagtagc tccggccctc tagctgtac tgtggccca 23460
gtgacctgac ttttccctcc ctccgcaggaa cgagatttag gagctgcggg ccgagatgct 23520
ggagatgcgg gacgtctata tggaggagga cgtgtatcag ctgcaggagc tgcgacagca 23580

-continued

gctggaccag gccagcaaga cctgccat cctgcagtac cggctgegca aagccgagcg 23640
ccgcagtctc cgtgccccc agaccggcca ggtggacggc gagcttatcc gtggctgga 23700
gcaggatgtc aaggtcagcc tgggctcggg tgcccttgtt gctggatgg gaccacaagt 23760
ctaagtgggg cccaggttgg ctttcatct catttcgtg tgtccatggc caagttctac 23820
cacctctctg ggcctcgtt tcctcatcta agtaggaatt cctggtaac ccacagggt 23880
tctatcaagg gtatccaaa atgatatcat aatatgaatc ctccctacta gcagcagcta 23940
ctgtttttt ttttccttt gagacagagt ttctcttgc ttgcccaggc tggaatgc 24000
tggcgcgatc tcagctact gcaactctg cttccggga tcaagcgatt tctttgcct 24060
cagcctccca agtagctggg attacaggca cctgccccca caccggctt attttgtat 24120
tttttagtaga gacagggttt caccatgttg gccagctgg tttcgaactc ctgcacctag 24180
gtgatctgcc caccttggcc tcccaaagtg ctggggttac aggcatgagc cactgcgcct 24240
ggctgagaag ctactgttta tttagtatca accaccaccc ggtgcttagcc tctaattgtc 24300
cccgacacgt tccatgggg ttttttttcc ttcttgcgtt tttagagaca aggtcttgc 24360
ctgttgccca ggtatcggc agttagtgc tcatagctca ctgtagcctc caccctctgg 24420
gctaccccca ctcagcttc ctgagtagct gggactacag gcacacacca ccacacctgg 24480
ctaatttttca tatatgttgt agagatgggg ttcattatgc ttgcccaggc tggctcaaa 24540
cttctggctt caagtatcc tcctgccttgc ctccatccaa gttgtggat tacaaggatg 24600
agccacggtg tcaggcagag gttccatatac tatgctaattg aaaagtagtg aaggcctcg 24660
gaccatccctg gctaatacag tggaaaccctg tctctactga aaacacaaaaa aattagccag 24720
gtgtgggtggc aggcatgtgt agtcccagct actcgggagg ctgaggcagg agaatggcgt 24780
gaacccacga ggccggagctt gcagtgcggcc aagatcgtgt caccactgca ctccagctcg 24840
ggcgacacggcg cgagactccg tctcaaaaaa aaaaaacaa agagaagtag tgaaggccta 24900
gagttggcttca aggctgggtt aaatcttggc tccatatac ctgaacttta ttatggtag 24960
caaagtcatc aatttcgtgt tcctcatctg taaaatgggt ttgcagcaat ctctatctca 25020
cagggttata tgaggattaa atgaaataat gccccgtggg cacagtggct catgcctgt 25080
atccccagcac tttggggaggc cgaggcgggc agatcacctg aggtcaggag ttgcagacca 25140
gcctggccaa catgggtgaaa cccatctct actaaaaataa cacaattag ccgagtgtgg 25200
tgggtccgc ctgtatccc agtactcag gggctgagac aggagaatcg cttgaaccccg 25260
ggaggcggag gttgcagtga gctgagatcg cgccactgca ctccagccag ggtgacacag 25320
tgagactccg tctcaaaaat aataataata ataataataa taataataat aataataata 25380
atgtgggtaa aatgccagca taaggactca aaaatgttag ctatcattat ttccatcc 25440
ttgatgaaga atctgagggtt cagagaaaag caatacctgc cgaaggtaa gggcgctca 25500
gtgtgttgtt ggcagagccc aggctcatga gcctccaaag cttccagttt ggcctactg 25560
ctctctgcctt caggaagcagc gggagggtt ggtgggggtt tgccagggtt gatgtcccc 25620
tgcctgtctt cggaaagctt ggtctgggtt tgagaactcc agggccccca ggcccggecc 25680
tctgcctca gaactcagcc cagttgtatc tggtgacacg aggagctggg atggggccct 25740
agggaggctt cctgcagggg tgacttctga gtgaaggccc aagggagaag tgggcctgag 25800
ccaggaaggg gaaaaagagga aggtggcaga agatggctcc tgtgaggacc caggggagcg 25860

-continued

gcaggatgct agtcagccac acacttacct ggggagaagt cccttctgtt ggaagacact 25920
ggcaaactca agccctatag ccttccagt ggaatataaa agaaaaacct ttgttaagg 25980
accaggaga gccatggttt caaaggccag cagaaccagg ggaggtaaac aggagtgcct 26040
gttccaggaa gctgctggcg cccctgtgg tcttgggtt gcacaatggg attgttagtgc 26100
tgtcctacac atgaggcaac caatggttgt gtctccgcca gaggtggagg ctcatgcctg 26160
taatcccaga atccccagcgc tttgggaggc tgaagtggga ggatcgctag aggccaata 26220
ttggaaatca ggctgggcaa catagcaaga ccctatctct aaaaaaaaaat aaaaaaaaaatt 26280
agcagaggct ggggggtgt gctcacacct gtaatcctag cactttggca ggccgaggcg 26340
gatggatcac ctgagggtcag gagttcgaga ccagcctggc caacatggg aaaccccg 26400
tctactaaaa atacaaaatg agccgggtgt ggtggtacac gcctgtactc ccagctactc 26460
aggaggctga ggcagatgaa tcgcttgaac ctgggaggc gaggttgca tgagctcaca 26520
ccattgcagt ccagcctggg tgacagagtg aaactctgtc taaaaaaaaaaaaaaaatt 26580
aaaaaaaaatt agctgggcgt ggtggcaagt gcctatagtc ccagctatat gggaggccaa 26640
ggccagagga ttgcttgcgt ccaggaggtt aaggtttcag tgagctatga ttgcagact 26700
gcactccagc ctggggcaca gagcgacaag gtctctctct aaaagaaaaaa aaagaatgag 26760
ttagtttcca agttcaaagc gtacctctac ttgacaggc agtccccggg catattcaga 26820
ttccaaatcac tgggttggga ggggtggggg gtgggtgggg aaggcagtag ggaggagatg 26880
gggagggggg gggaaatagct ggagaagctc agaggaatct gcgtttgggc atatagtgg 26940
ggatattctgg gaggacaca aaaaaatgtc agacttcaga aagatgagga agaggcggag 27000
gagaggctgg gacctggagc tagctgccgg ctgcggcgcc tccgtgggtg gagctagagc 27060
agcggccggcg gcccggggg ggcgggtgtg gggagccggt gtgtggctct ccggagtcg 27120
ggcggagagg agggggcggt tctgttgggg tttgaattag ccaatcagca tccctcttc 27180
cctccctcc ctctctccc cacgtgeag ctggagcgct gagctcatct gggggcctgg 27240
ggcctgagct ggtggacac ctctggccac caccctgagc cccatgtgtc ggggtcggtt 27300
tgggggggtt tgggaagtta tcagaatgct gcttccctg gcacgtctt ggaggcctt 27360
atggacccga gtggggaggg ggcggggcgct gcctaggagg gtacctctgg gacttcccc 27420
acttgcctcg ccagctccgt ggtccatgcc ctgacttgcc cctaccgctt ggtgcctcat 27480
tttcccttc tgtaaatgt ggctttgac cagctgaatc ctcaggcctt ttgagctct 27540
gtggtttat ggagcagagg ggtattgggt ggggtgggtg gggggcattt aataaaacca 27600
caggcctcag tgcggggggg caaccccccag ctctccctct ctgctggtcc tcaccgtt 27660
caggacctac agaacctcg tggggccctt gcccagcatg aggaggccaa ggtcacgggg 27720
cgagagactg ctgtggccacg tttgtgtgtt ggcctgggt aagttccctt acatctctg 27780
gcttgcctcc tcatggaaaa tggagataaa atgagtcctt ccctcttagag gttttgtga 27840
ggataaaatg aaataattca agtgaagaga gtattaaata ctaggctgg aaataaaacat 27900
gtattgggtt tggttttttt tttttttttt ttttaaggc aggggtctcac tctgtcaccc 27960
aggctggagt gcagtggtgt gatctcagca acctctgcct cctgggtcc agcaattctc 28020
ctacctcagc ctcccaagta gctgggattt caggcatgcg ccaccatgtc cagctaattt 28080
tcgtatTTT agtagagatg gggttctct gtgttggcca ggctggtctc taactcctga 28140

-continued

cctcaggtga tctacccacg tcggcctccc aaagtgcgtt gattacaggt gtgagccacc 28200
atgcctggtc tgggttggtt ttttttattt ctgagaagaa ctcacacacg aagtataagt 28260
tcttcgtggc aatacatcag taggactact agtagtagta ggcttaggtt actgggttgtt 28320
agtagtagta gacttagtagt ctggggcgaa gggttaggcca ggatctcagc ctcccagggt 28380
ggataaaaaat caccaggagg gcttgcac aacatacatt cccaggaccg gtgaggagct 28440
gagaccggat acaaggaggat caggtccagg gaatctggta tctaaactct ggtctaaagt 28500
gactctgtg ggaagccaag gacacaccag gacacgccc ctttcctcag atacagatgt 28560
tgctggagaa aatgttctgg ctttgggtt cattcctagg tgacctgtgc aggttacetc 28620
ccttccttgc gcctcacttt cctcatctgt agagtgggg tgacattcct accctctgg 28680
ggttttagg aggattcagc cacacggagt ccagaatgct gcctggccag gggatgctct 28740
tccttttttc tttttttttt ttttttttagt atgaagtctt gctctgtcac ccaggctgaa 28800
gtacagtagc gtgatcgctg ctcactgcaaa cttccaccc tcgggttcaa gcaattctcc 28860
tgcctcagcc ttcccgatgt ctgggattac aggcattgcca ccataccacg ctaatttttg 28920
tattttttta gttagagacgg gggttccacgg tggttgaccgg tctggctctg aactcctgac 28980
ctcaagtgtat ctgcggccct cggccctccca aaggggccaga attacaggca tgagccacca 29040
tgcccgccct tgtttttttg tttttttttt tcttttttgc aacatagtct ctctttgtca 29100
ccaggctgc agtacgggtgg ctgatcacag ctcactgcag ccttgaccc tcgggtctcc 29160
aagggtatccct cctgtctcag cttectgagt agctgatact gcaggcacac accatcatgc 29220
cctgctaatt tttttttttt tttttttttt ttttttttttgc tcaggctgg 29280
cttgaactcc tggcctcaag cagtcttccct acctcgaccc cccaaagtat tgaaattaca 29340
ggcgtcagcc accatgccccg gccaggagct gcttttcgt gagagctccc atccctctcc 29400
ctgaggatca ggtatgacaga gcccaggag ccgagggggcc ttagtgcctgg ggagcagggg 29460
ctggaggctg ctagttgcac ttggcaggag ccaggcccta gcgtgtaaa cagtgcaccc 29520
atggctgagt gcattttttt atggttttagt gatttgcctt ggtttccggaa agccgcagct 29580
gcccgcctc ctttagtatac tcagcccccc taccttgagg ttagtggaa ggatcaaattg 29640
ggatagacac gtggggccct tcaaattgcct gacacagtaa gcccactgtt ggggtttatt 29700
attaataactt gtatttttgc tggtaatgtt atgtttttt acgcattggg cctatacgcc 29760
aaaatcagcc tgagagtgatg accttagatg tgacccactc gatttcaaac cttagttcta 29820
tagccttcct agctgtgtga ccttggaaag tgaccccttcc ttcttgagcc tcagaatggg 29880
gataataatc aaagtgcctt ccaacggat gaaaagcagc tggtgagcac ccggggatgt 29940
ttccagctga ttctgttaat cacttggatgg ggccgggggg gcccaggacc actgaggctc 30000
actgaggagc agggacaagc ccaggccaca gggctcctgc ctcccaactt ggctgcctgg 30060
agttgtgatt tttttttttc tttttcttagt acaaagtatt gctctgtcac ccaggctgaa 30120
gtgcgtggc gcaatctcag ctcactgcaaa cttccgtctc ctgggttcaa gcaattctcc 30180
tgcctcagct tctcaagaag ctgggattat aggcatgtgc caccacgcct ggcttaattt 30240
tgtttttta gttagagacag caaagacggg gttttaccat gttggccagg ctgggtcctga 30300
actcctgacc tcaagtgtatc tgcccgcttc agcctcccaa agtgtggaa ttataggcgt 30360
gagtcatcgt gcctggccctg tgctttttt tgatacagca gccatcaggaa aagggagta 30420

-continued

```

aataggaaag ggattccatg gagaatgggg cagaaggaa ctgggaaatc atggagagga 30480
tgggagggga taaaatccat agagttgttc atgaacccc gaggatccag aaatccacat 30540
ggatttgac ttcttgcgtgg ggactggtag cctcccccagg tgaggacat ggtggagatg 30600
ggcatcccc gcaggtggag aggagctggc ccagatggga gctcattga gagaggagcc 30660
ctcaccccaag ccaggggtgc agcccttct ggatagtggg tgggggtggc atttgagggc 30720
agcctcttca gcccctgegag ggcaggtacc actccccatg tatattcagt cattcactca 30780
ttcatctccc tccctttca ctttgtggac ctttgattt atttgttcat ccaacaagca 30840
tttatttaggc acttagtgtg tgcctgacgc tgcgtgctg ctggacaca tagatatgaa 30900
tgaggccaag tatgtctgct gccttgcattt ctgtacccca ccactcctgc ctccagtgcc 30960
taagcaggat gagtgagtga gtgagttgt gagtgagtat gagtgagttt atgaatggaa 31020
tggtgatcag gacaggggct tactgggtga cctaacagcc ccaaactggg gaccatgttt 31080
aaataactta acctgggtca catagccagt aaatgctgga actgggttgc aatcttagtt 31140
tcctatctgc aaagcccttg ttctgatact gaggggttgc tctcccaact tcgtggatga 31200
ggaaggctcg gtaccccgag aggctgtggg aggagcctca tggggccgcct cctggctgtc 31260
ctacgagatc ctcgggcaga aacaggcagt ggcagtgccag ggagggagag gccatgcctg 31320
gtttcccttc tggtcgttgc ttccctggc ctgcctcata cattgccaag gggccgtgc 31380
cgagctgggc ttccggccag gcagaggta gagccaggga gggcccttgc ccctctggc 31440
aaggaggagg cacaacagtc atttccgaga ccagcaggca tgggttgctg ggctctctgc 31500
cccacatgga gttccctgtc gcgaaacctg ggcacccctg ggaaaccttg ggctatactg 31560
gcccctcagc caaacccctg gggctccag cccagctaa cccacactgg ctttgcgtga 31620
ccctgagttga gtcacttctg ctctgagtc cagttccctt gtaaaatggg gacaggattt 31680
gggtgaagat ggggtgagttt ggcagtcgtt aggtattcac caaacagtag ttcccaatgt 31740
acatagagat ggagtctgg tgcatttcc cagccagaa gggcccttgc ttccatccc 31800
tgccaggccctc cttccctggc ttccctcaac cagccagccag gggacagaac agagggtgtc 31860
taccaggaat ggggtgtgt tgctacagag tccctggacca ggatcagaga ttgcagctcc 31920
tctagttct ttctgttaggg atgacttgga catctgccc cccttgcata ttttggccct 31980
gttctacatc ccagacaact agatatacgat ctcagggtca cactgtgacc tccagcgatc 32040
atgccttct ctgggcttca gttccctggc gcagcagctaa ccagtcaccc agaactcact 32100
gtgtgcagggc cccaggccaa ggctttgtg taatagctgg ttcagttatc ctgtcggcc 32160
tgtaagggtt accctgttat cactgcgttc atttatagt tgaggagact gaggtacaga 32220
gaggaaaatt gccttgcaca ggactacatg ctggtcactg gcagagatgg gattcagtgg 32280
tctaggccctc atcttctctg ggttccact gcaagtcata cccctctctg ttttggccct 32340
ctttgtttt ttttggccctt aggagacagg gtctcgatc gttgcctagg ctggagttga 32400
gtggcacaat catagtcac tgcacacccg aactccaggg cccaggtat cttccatgt 32460
tggccctctg agtagttggg cctgttaggt tggtccacca tgccaaatc actttttgg 32520
gtttttgttta gagggtgggtt cttgttatgt tcccccaggca catctcaat tccctggcc 32580
aagcgatccc cctgcctcg cctctcaaaag tggggatcc caggtgtgag ctaccgtgtc 32640
cagccctca aagtgttggg attacaggtg tgagctactg taccctgttc ctcagtc 32700

```

-continued

tcccttaag atgagttct ggcctcca caggtttg tgaggactgg gcaagtgg 32760
gctggagag gcttggaaag tgtgaagagc ccaggttaggc aggcacagtg gctcatgc 32820
gtaatctcag ctactcgga agctgaggca ggaggatccc ttgaggccac aagttcaaga 32880
ctctctgga ccacatagtg agacccgtt tcttttatt tatttattt tagacaaa 32940
tttgctcta ttgcccaggc tggagtgcag tggcatgatt ttggctacc gcagcctctg 33000
cctccctggt tcaagcaatt ctccgttc agcctccaa agtgctggg ttacaggcac 33060
ctgccaccac gcctggctaa tttttgtat ttttagtaga gatggggttt caccatgtt 33120
gccagtctgg tctcgaactc ttgacctgag gtgatccacc tgcgtcgacc tcccaaagt 33180
ctgggattac aggtgtgagc cactgcgccc ggctgagacc ccgtgtctac ataaaaatgc 33240
aaaaatttgc cagtttggt tatgcatacc ttagtccca gctactcagg aggctgagg 33300
gagaggattt cttgagccct ggagttcaag cctgcagtga gacatgattt caccactgca 33360
ctccagccctg ggagacatag tgagaccctg tctcaaaaaa aaaaaaaaaa aaaaagcct 33420
caggtaaatg aaaaggcatt cctgagaaca gggtcttccc agctttctgg ggctgtgt 33480
tgatacagat gctgagagct tgccgtaccc aggtttgggt tttggttcta ctgttcctt 33540
catgagtttcc ttggctctt tgcgtcgat tttgctcata tgcagcgtgg cagtgtatgc 33600
ataggaacct gcctgtctt agggttgccc tgctgtgaaa gtcgggtggt cagtggact 33660
tgcagtgggg gctctctca gcctgctac atccttcctt ctcccccccc tgcacagg 33720
tctaaggaca tctccatgcg gctgcataag gagctcgagg tggtgagaa gaaacggcgc 33780
cggtggagg aggagaacga agagcttcgt cagcggtca tcgagactga gctggctaa 33840
caggtgctgc agacggagct ggagcgaccg agagaggtga ggacctcatg catccagcag 33900
ggccagcccta gggcaggtgg gcacatgcaaa gccccccaa gccagtgtct ctgtgagccc 33960
ccgcttggaa gaaatggca gcttatgcctt ccctggcaag gttgtttatt cattcattcc 34020
tccaacatgc atttcctgag cacctactgt gtaaattgtg tgctggacac taaggttaca 34080
atggtaaca aacacacccc agttcctgtc ctcaggagc tcacggggta atgagagaga 34140
gagtcattca tcaaagaacc acagccgctg tgggaggatt aactgagaaa acatgcaaca 34200
atgtcaggca tggggctcag ccctctgtgc ctcagttcc ttacctatac aatggggcag 34260
tggtgccctg agtcattgtg aggaccaaca ggaataacgc atgtggagt gctgcaagag 34320
actgcgagat cctgcccattt ggtgtgaagt cagtcaggg gctgggtctc aggccagtgt 34380
ctctggggta ctgggtcttc ccgaatgtga ccctggagaca cattcctgtc tgggttttt 34440
catgccagtt cctccctgg cagaccttct cctctgtt gttgtgggt atacagtgt 34500
ttatgcggcc tggatgacag tttatctcaa acatttcat gctaacaagg ttgcgtatcat 34560
gcccattttt tggatgagca aacagccctt gagagagagg ggggacaac ctcaaggcca 34620
tgcaaggcctc agggggttct gcagaatgcc ggctagaaac catcagccata ggtgtccctt 34680
aaaggctgga atccccccat tccagggttc tgggtccaa atggcctggg gtggcttctg 34740
gccctagagg cagagccagg ccttagctct ctgcgtgtc atgagatatg aggatcgtgc 34800
attcttcttag ggcacactcc tcgtgtcagg cgctgcctgc cttccctccc cgagggcag 34860
cctgccagtg tccttgact gagacaccac agcccaaggc cagaacattc ctccagccat 34920
ccttccact gggtgtgtt gttgtatgtc ccaatgcct ctcagtgtc agggagagg 34980

-continued

ggcgtttgc gctaagggtg gctagcccc tcccttcatt cagctgccca ccccatggc 35040
ctggccata acacggagca ccatcttcac accgtttgc tgctgcaacc cttgatagt 35100
aaaaggggag acaattctg aaagcattt aaggcacgg tacaagtaat gaaactgagg 35160
ctcaggagga acagcttgcc tgagttcacc cagcctctcc tttagcagaga tgtggctgga 35220
accctgaago cacttcccct ctctaagtct cagtttttt atctgtgaca tcaggataag 35280
caagcaaaca aacaagcaaa gcctgcaegg tggctcacac ctataatccc agcactttgg 35340
gaggccaagg tggcagatc tcttgcgcct aggagttcga gaccgcctg ggcaacatag 35400
tgaaaactctg tctctataaa gaatgcaaaa attatccagg tgtgggtggt cctgcctgt 35460
gtcccagcta ctttaggaagt ggaggctgca gtgaggtgag atcgcaccac tgcactccag 35520
cctgggagac agagtggagac cctatctcca aaaaaaaaaa aaaaaaaaaa aagggggggtc 35580
caggcacagt gggtcacgccc tgtaatccc gcactttggg aggctgaggt gggtggatca 35640
cctgaggtta ggagttcaag accagcctgg ccaacatggc gaaaccccat gtctactaa 35700
aatacaaaaa aaaatttagt gggtgtggg gggggcacct gtaatcccag gtacttggg 35760
ggctgaggca ggagcatctc ttgaaacctgg gaggtggagg ttgcagtgag ccgagattgc 35820
gccactgcat tcattccagc ctggcgaca aaagtggaaac tccatctcaa ataaataaat 35880
aaataaataa ataaataaat aaataaagca ggatgataac tgtacgcagc ctatgaggtt 35940
gtttaggatc attaataaga ttatgttagcc atagaataact atgaagaat taaaaagaat 36000
aagatagtagt ggtagttcct caaaaattaa acatagaaca atcatatgat ccagcaatct 36060
cacttctagg tataatccaa aagaattgaa agcagagact caaagagata tttacacacc 36120
cgtgttcata gcagcataac tcacaatagc cagagggtag aagcaacccca aatgttcatc 36180
aacagaggaa tggataaaca aagtgttagta cacatgtata atggaaaact attcaacctt 36240
aagaaggaat gaggctgggt gtgggtggc atgcttgc ccccagcatt ctgggaggca 36300
gaggtgggca gattgggtgg ggcttaggatg tcaagaccag cctgagcaac atggcaaaac 36360
cccgctctca caaaaaatac gaaatttgc caggcatctt tgcgtatgcc tgcgtatccc 36420
gctactaggg aggttggatg gggaggatca cttgagccca agaggttgag gctgcagtga 36480
gctatgatgg tgccactgga ctctagcctg ggctatagag tgagaccctg catcgagagg 36540
gagaggagga aggagaggga ggggggaggg ggaggggggaa ggggggggaa gagggggagg 36600
ggggaggaga gggggggggg gaggggggggg gggggaaacag aaaagaaaag accagcactt 36660
tgggaagccca aggcaggccgg atcatgaggt caggagatca agaccatccct ggctaaacac 36720
gtgaaacccc gtctctacta aaaatacaaa aaaaattag ccaggagtga tggcgggcac 36780
ctgttagtccc agctactcag gaggctgagg caggagaatg gcatgaaccc cggaggtgaa 36840
gcttgcagtgc agtgcgatggat gtgcactgc actccagcct gggcgacaga acgggactcc 36900
gtctcaaaga aaaaaaaaaa agaaaagaaa gtatggaa tttatggaa aaatggaa 36960
aagaggtaaa gcttgaaag agccctttgt ggtgcctggt acagagtagg tgctcaagaa 37020
acactggttc acatcccttt gcagctggtg tcagtgaggc tggcttagct ctccctcgct 37080
gcccctgtggt gtctcaggct tgcacagg aggcgggagc ctgtgtttat tatttgcgt 37140
ttatccctctc ctcctccgt ctcttctcg acatggttc tatggagacg agggccctg 37200
agcaattctt ctgcccagagc ctcagtcaca tcatcaagat ctgcccgtgc catccctgaca 37260

-continued

ggctgccgag gtgggtgtgt caccggctg ggaacaggag ccctggct gagctgtgt 37320
ggttgccatg gagacagctc tccgcgttagag agaagggagg tggcagaccc tggccgagg 37380
ctgcgggtggc actgttgcctc agctaccctc tcttccagtg tcacaccct ccttgtgttt 37440
taggcagaaa gccaatacga gtgttgcgt gatttattat ggtatagctg ctgtgtttga 37500
gcctcaactc tgtgccaggc accacagaag cttgaccac accagcacat tgtttaaccc 37560
tccaacagcc ttcccacgga catggtgtgg ttgttattt ttataggtgc cactgtttgt 37620
tggcaatttt aaaatcatga tttaattccc atatcataca actcaccatt taaaagtgt 37680
cagtttgggc tgggcacagt ggctcacacc tctaattccca gcactttggg aggccaaggc 37740
gggtggatca cctgaggtaa ggagttcgag accagctga ccaacatggt gaaacccac 37800
ttctactaca aataaaaaaa tttagccaagt gtgggtgtgg ggcctgtaa tcccagctac 37860
ctgggagggtt gaaataggag gagaattgtt tgaacccggg agctggaggt tgcaagtggc 37920
cgacatccca tcatttcact ccagcctagg caacaaagcg agactctgtc taaaaaaaa 37980
aaaaaaaaagtgt acaattcaga ccagacacag cactttggg ggcggaggca gcaggattgc 38040
ttgagcccaag gagtttggaga ccagcctggg caatataattt agacccttgc ctctacaaaa 38100
atttaaaaaaaa tttagctggac ctgtatcacac aagcctgtgg ttccagctac ccaagaggtt 38160
gaggtgggag aattgttgcgaa gcctggaaa cggaggttgc agtgagccaa gatcacgcca 38220
gtgtactcca gcctgggtga cagagcgaga ccctgtctga aaaaaaaaaat taaaataaaa 38280
acaacaaaaa ataaaaagtgt ataatttagt gatttttagt atattcatga tgtttggtag 38340
cgtcaccatt atctaattcc agaacatttc catggccca aaaaatggc tcatgcctat 38400
gaacagtcac tccccactcc ccctttccct cttgccccca gcaaccactc atccgcttc 38460
tgtctctggc ttgcctgtt ctggatgtt catccaaatg gagttttgg tgactggct 38520
cttttactta gcatcagggtt ttcaagggtt atccatccat gttgaagaat gtagcaggac 38580
ctcatttcctc tttatggctg agtaatcatc cattgtgtgg gtatatgacc acatttagatt 38640
tgtccattta tctgggttgtt tttggctact atgaataatg ctgctggaa cattcatgtg 38700
cacgtttctg agattctgtt tttagttctt tggtagatc ccagggagta gaactgcagg 38760
atgttagacgt gactggggaa agacaaagat gggattcaaa cccaggactc cacatcttc 38820
aaacccgcac tgactttctt ggtgtgtcc ctatgtctt tgcccccacc ttgtcttc 38880
aggaccctgt tggtagttca gaaacctggaa agaggctgcc cactcctgtg actgtttcc 38940
tccctggcac agccttcata tccggaaattc ctacctgggtt ggcttagaaag gggctgtata 39000
ttcaaagccc ggttctgcct ttttccaga ctggcactt acttgcaagt gactgtacc 39060
ttgcggaaat gatataactg ctctagcgct tagtttattt attttattat tattttggg 39120
ggagacaaga gttttgtgtt gtcactcagg cttagtgcgtt gtggcacat catggttcat 39180
tgcaagctca aacttccggg ctcaagagat cctcctgcct cagcatccca agttgttggg 39240
actataggcg cacggccacca cacctggata atttcgattt ttggaaagag atggggcttc 39300
tcactatgtt gcccaggctg gtctcaaact cttggctca agctatctgc cttctcagc 39360
ttcccaaagt gctggattt caggcatgag ccategtacc cagccttagtt tgttataaa 39420
atggaaatat tatccctact tctcagaggtt gtttgcgtaaa attaaacaag acaagcaagt 39480
aaagtgccta acacagacta aggacttattt attttcattt gtgtcacaac caccgtgagg 39540

-continued

gctgegggtc tggagaagca tttgtcttag ttaaataaat acatttctgc ttccctctggc 39600
aaggagctct gtgtcttagag ctgccagttg cccggaaag gagcagctac gttgccaatc 39660
ttcacgtgca gttggcttca tctgtccttg tgacatttga tttggggcca atgtttaaaa 39720
caactttgtt tttctcccc aagtgtatctc tccctaatac agagccacat ggccctcttc 39780
tgtggccatt tcagccactc aggaggaaga atgaatgcct aaatcttga ggtacacccc 39840
caaccccaag ttcccttctag aatctacagt aggccggca cagtggctca cgccctgtat 39900
cccaacactt tgggaggtca aggccgggtgg atcacttggg ttccaggatcg cgagaccagg 39960
ctgggcaaca tggcgaaatc ccatttctac aaaaaataca aaaaatttgc cagatatgg 40020
ggtgacacacc tgttagtctcg gctactcagg aggctgaggc atgagaattt cttgaacctg 40080
ggaggccggag gttgcagtga gccgagatca tgccactgca ctccagcctg ggcgacagag 40140
tgagaccctg tctcaaaaaa aaaaaaaaaa gaatcttagag taaatttgtt acctgaagca 40200
cagtctgtgc tttcagttag acctttccaa acatctccctc cagatttggt gaaaatctgc 40260
cctgtgttt tggcagaaga ttcattttgtt tccttggaaag gttggctttt gtttcttggc 40320
caaagtgggg caggaaggtg tttgagtgca gaggccctc cagagctgtg ctcaggcagc 40380
tctctccctgg gctgaaggaa agacaaggga ggacatagct caggcttctc cttagaaggct 40440
gtagacaggg agatgttttag aggtaccaca ctggctgctc aaagatggcc ttgtgagttg 40500
ctggcgagtg gggaaagcccc tggcatgc caggccacag agtttgcctt cagggctggg 40560
catgtcccca ggccttgcgg ctgcatctc gctggggctt aggcacccgc ctgccttgc 40620
ctgcccctcc aaggcccaaga gccagaacat ctgccttgc ggcagagact ggcttcatcc 40680
ctgggggctc cacaaagggg ctctgacagg ccctggctttt cttgggggtt gcagccccc 40740
agccacctgg ccagctgtg ccgtgcagag gggaggaagg tctttgtgca ctctgagccc 40800
gccttgcctt ccctccagaa tggcagggtt gtgtgggagcc acatggct tagataagg 40860
gggggaagtg tcaggtctca ccagaaaagg gtggaaactcg tctctctgc ccccaaccct 40920
gctttctggt gagattcaag ccactcagct cattcacacc ttttctctgc ctgtggaaa 40980
ggtgttttgtt gacttcttgg ggaaagtata ttttaaaaac cttttataat gactttccc 41040
ttccccccag caggttagtag tatgtgctta gcagaaaaag attagaagat tcgaataagc 41100
catagggaaat ttctgtgtgg atggggaggt gcataatactt ccaataagtt tttcttatgt 41160
tacatatatg tgtatTTAA tgtaaacgtg aatatgtttt ttttctttt tggagacaag 41220
gtctggctct gtggccagg ctggagtgc gggacgtgtat cttggctcaa tgtaacctcc 41280
accttgcggg ctcaagcaat tctccaccc cagccctcca agtagctagg actacagaca 41340
catgccacaa caccggcta agaatatgca ttttgcgtt tttgtttttt ttcttgc 41400
tgagacagag tttctctctt ttggccagg ctacagtgc gtgggtgtat ctggctc 41460
tgcaacctct gcctctggg ttcaagtgtat cctcctgcct cagccctcccg agtagctggg 41520
attacaggca tgacccacca tgccggcta atttttttt ttgtatTTTT agtagagacg 41580
gggtttctcc atgttggta ggctggctt gaactccga cctcaggtga tccacccacc 41640
tcggcctccc aaagtgtgg gattataggc gtgagccact gcatccagcc agaatatgt 41700
tttttaatc taaaaaagta atataaactc agtgaagcaa aaatatcatt agaaaagcac 41760
acagatgcat gccttagcata cctccaaagga caagctgcag tgctctgggtt acgttgggtt 41820

-continued

ccttggggcg ggataactgg gtgactgggt tcactgtatt cttctatgtc cttttgaat 41880
gttgaaccat gcaactttgt tatctattca aaaataatag aaataataaa acccataaat 41940
taaagaagaa agccttcct aactgcctcc ttcaattcta ctccctcagcc tgaccagca 42000
tcagcagttt gctttgtatc tttccaggc tttctctagt ctcatacaat gcaggtttct 42060
atatattttt aaactttttt ttttttttga gacagagtct tactcttttgc cccaggctgg 42120
agtgcagtgg tgtgatctcg gctcaactgca accactgcct catgggttca accgatttctc 42180
ctgccccagt ctcccttagta gctgggatta caggtgcctg ccaccacacc tggctaattt 42240
ttgtatTTTT agtagagaca gggtttcaact atggtggcca ggctggctc aaactcctgg 42300
cctcagggtga tccaaactgcc tcagcctccc caaaagctgg aattacagggc atgagccacc 42360
acacccagcc atttcttct tctttctgtg gctgagtaaa tattctgttg tgtggataga 42420
ccacatTTTT ttaatccatt ctaaagttgg tggacatatt tgggttatag gctttgtttt 42480
gctggtaact tctttccctt aacagagtgt ctgagagact ttaccatgtc agtgaatctc 42540
atgatccctt agtactttga atggcttttag gtgatctcat tcattcacat gtgatgcctg 42600
accccttggg acctatgcag acccctggct ctccctctgt tcacttcact tccttcctcc 42660
agggagtcgtc cagtcattgg tggggctgtc gtcacctaatt ccatggccca gtgtataaaa 42720
tgagaaaaatg gaatttcagg gagaggaaag gggtcaccca gggatcataa actcagcgtt 42780
gcaaggggcc cagagttcaa ctcccaagtcg aggccacago ccttcccact cagttgtaa 42840
taggtgggcc tagatccttc catagcctga tggctctgac ccacagcacc cccttgcatt 42900
ctatacgggg gtctgtttct gtgtggccctg tccaggggccc cactggctgc ctttgtctgc 42960
tgcctccctgg ctcttggaa aggactcggtt ttctgacccca tagccctgtt tcctccctttt 43020
cccttccagc ttttggaaagc tcccagggtca gcacttctca aataactggct cctggctttt 43080
ggcacccagc gtctgaaata tcagcctgtc cttccctgc tcatttgcatt ttcaaaagggt 43140
ctcttagga ttgttatttttttgc ctcttaagata ttctgtactt aggagtccat ctggagctct 43200
ccttgcctgg cacagaggat ccgacactgca cttcagaatc acctgaaagc atttggaaaa 43260
tactgtatgtc tcagacactac ctcagaccac cagttttttt ttttttttttgc 43320
agtaattctg atgtgtacca gggttggaaac cactgtgtcg gctacactct ggcctcatc 43380
agagaagaca cattttctcc taccaactct ttgaaatttg gtcaagagct atgtcaagaa 43440
ctgggtcccgag gtcaagtcgtc ggtgtactgt gagaccttgg acaagtcctt gcccctctgt 43500
cttgggtttca gaaaaatggg actcagattt gatcagtgcc tgcactttc agggatgggg 43560
tgtgagaggt gggggatggc gcaagcatgt atccaaatca cctggggccca ttttcaaaaa 43620
tataccagcc caccttggaaa agtcaaaaaca acctaagcaa ggtttatac atacaccacc 43680
accaggaaatc accccttcctc tcttggggagg aagcgctcaa aagaatcgatc catctaaattt 43740
aatgacttct ttaaacatgg agaaatttca gtaattgtta ttgttaacatt tcctccaatt 43800
ccctctgttag ggcttgccttgcctt atttggcattt cttgcctcagca attatgtacac ctttttttttgc 43860
ctctatagct tcaccaatag aatatgtcgat cagatatttgc gatTTTGCCTC aatctgtatgt 43920
gtgagaaacg gtatctcgat gtaattttaa tttgcattta tctaacaggaa gtgggggttgc 43980
atatcatcat gtgccttgat gtcttcatgt gccatgtatcc atttgcctttt cttttccctgt 44040
gaactgtttt tcatatctct agcctatTTTT cctacaggcgc tgctggcctt ttaatgtctc 44100

-continued

gtttgtgaaa ctcttccat attaggata tcaaccctt gttgggttg tacattgcag 44160
atatttttc agagtttgc atttatgttt ttactttct taagggtttt ttatttccat 44220
gcagaatttg cttgtataat caaatgtatc tatttttcc cctattgttt ctggatttt 44280
aacatcttat tttttgagg tggagtcttg ctatgttgc taggctagtc tcaaactcct 44340
caatgcaago aatcctctca cctcagccctc ccaaagtact gggattacag gcatgagcca 44400
ctatgccaga ccagcttctg gattttgaa tcactgtttg ccatcagtct cttatgacca 44460
tttctaagcc tcatgtgtaa ttacgatttg ttatattcat ttattgttg ttgtttgtt 44520
tttgagacag ggtcttgcata tggtggccag gctggctca aactcttgc ctcaggcaat 44580
cctcctgtct tggcctccca aagtgttggg attacaggca tgaactacca catccagcct 44640
ccatattcat tttatccatt agtattcatt ctgaattaca gttcccacta gatgtcagg 44700
tcatgaagac aggattttct ctgagatgga gatttgcattg cagaggtttta tggaagagtc 44760
tccttggaa caacaccaca gggagtgagg cagagctgga agctgaatta caccagtat 44820
ctcaggcgat cccatgggaa gctctggagc taggatggcc ctccctgttag agctgaggca 44880
aggggtcatg gaccatcccc tggatgtggg ctccatccac ggatagcacc tgcccttgg 44940
caaggtggtt ctccctcaacc aatggctgtc gctgggagag gtactcagct ttgaggtctc 45000
agctctggaa atgagtgctc cagttccgaa ggggaatctg ggcaaagcac cacaacatcc 45060
tccctgggtc actgctgtgt tctctacacc tacctacctg gcgcttgcata ggtgctcagg 45120
aaatcctttt atgaaagaca aatgagaaca gaaacacago tttccaagtc ttgcgcgagg 45180
ccaactgggt cacacatggg agcaggctgg aggggaggag ccctgagcag gtgttttag 45240
aatcatcttc tccaatgctt ctggcctatc tcccattgtc gctcaactggg tcatctggag 45300
ggctccctttt actcaacagt cagtgccatc agggtggggt ggcctggggta ctggggccct 45360
ccctggaccc tgggcactgt gacagcttc tgcatgtca tcattctgtg agcttcag 45420
cctgactttt tttttttttt ttaattaaat ctcacttcag cattccttga agaaaagagg 45480
aacccgctcc ctggggagg ccgataagaa gactttggtg caggttaattc ccagcagccc 45540
ccgttaggaag gtttattct gaagacctga gctgactct aagtggctg gggatcccag 45600
attctgcctt gtcctccct tgctcatttt cttttttttt tttttttttt tttttttttt 45660
tttgagacgg agtcacactc tcttgcctcag gctgcgtgc agtggcaca tctcagctca 45720
ctgcaaccc taccctccag gttcaagcaa ttttcatgca tcagccccc gagtagctgg 45780
gactacagggc gtgcaccacc acacctggct aattttgtt ttttttagtag agacagggtt 45840
tcaccegtgtt gaccaggctg gtcttgcact cctggcctca agcaatccac ctgcctcggc 45900
ctccccaaagt tctgggattt caggtgttag ccactgcgc cggccccctt gtcattttc 45960
tttagttcac tgcccactt atttccaaa accacatctt gggccatgg ttcttagatg 46020
cttagatgtc cgctccccac cccagccagg aacttagcaa gcagcactac agtcaaatcc 46080
aaaggcttag agttggagac ccaggcttgc atcttgcctc tgccacttgc ctgcattgt 46140
attggagcga gtcattttga gtttctgggc tccatttctt atctatccct gtcttgggg 46200
ttcattggaa gaattggatg aagtactata tgcagcaaag gcccataacc gttagctct 46260
agtggatgg gattgtcaga gtgcaggcc cttttcatcc ttagcttatttta taatccctgc 46320
tgtggctctg ggagggaggt aatgttaccg ccccatatata ctgatgaaac tgttagtggca 46380

-continued

```

gcagttgctc atcagtatgt cacctgttcc tgtactccca ggctatgtgt ggcgc(ccat 46440
ccttgcgttc atttcaacat ctacccctct ggttctcccc cggaccctta gctcctttag 46500
ggtctggctg tggctgtattc atccccgtgt tcccagaca tagtgc(ccc caaaatgcca 46560
tggctggta c tgagactccc agcactaacc ttccctgggtt gtgtctgtct ctgtggccc 46620
aggaggacag tgcagaccc aagtgcagt tgcaacttgc aaaggaggag tcagccctca 46680
tgtgcagaaga gctcaactaag cttgccaagg agaatgacag catgaaggag gagctgctg 46740
agtaccgctc gctctatggg gacctggaca gcgcgcgttc agccgaggag ctggccatg 46800
ccccccactc gcgggagacc gagctgaagg tgcacccgtaa gctgggtggag gaggaaagcca 46860
acctgctgag ccggcccatc gtggagctgg aggtggagaa ccgaggccctg cgggctgaga 46920
tggacacat gaaggatcat ggaggtggct gtgggggtcc tgaggcacgc ctggcccttc 46980
ccgcgcgtggg tggcgagag tgcggggaga gcttggcaga gctgcggcga cacctgcagt 47040
ttgtcgaaga ggaggcccgag ctgcgtcgcc gctccctctc cgagctcgag gaccagaaca 47100
agctgctgct gaacgagctg gccaagttcc gctcgagca cgagctggac gtggcgctgt 47160
cgaggacag ttgttctgtc ctca(cgaac cttcacagga ggagctggcg gcccggcaac 47220
tgcagatcgg cgagctcagc ggcacggta agaagctgca gtacgagaac cgcgtgc(tcc 47280
tctccaacctt ccagcgctgt gacccgcct cctgccagag tacgcggccc atgcgtggaga 47340
cgacgcggca ggcggggac tctgcggcgt gtgtgcgttc tccctggg(g gagacacac 47400
agtcccatgc ggtccgactc tgcagagcca gggaggccga ggtgctgcct gggctgagag 47460
agcaggccgc cctggctagt aaggccatcg atgtcctgtt ggtgatgcc aatggcttca 47520
cggtggcctt ccggctgtgt ctggacaacg agtgcgtgtc ctccggctg catgaggccc 47580
ccgacaacac cgaggggcccc agggacacca agctcatcca tgccatcctg gtgcgcctga 47640
gcgtgcgtca gcaggagctg aatgccttca cgcggaaaggc agatgcgtc ctcgggtgt 47700
ctgtcaagga acagcaggag tccttctcat cactgc(ccc cttgggtcc cagggctct 47760
ctaaggagat tcttctggca aaagacccgt gctcagactt tcaggtaaagg tgccctcatgc 47820
acagatccca ttatttat(ttttcccttg ttttggggat ttttcttga gatagggtct 47880
ctccccgtg gccaggctga agtgcagtgg tgcaatctca gttcaactgca gcctcgatata 47940
cctggctct agccattctc cccccctca tctccgtagt acctggact acaagtgcgc 48000
accaccacac ccagctaatt tttat(tttt ttttggggat tttgagacgg agtctcgctc 48060
tgtcacccag gctggaaatgc agtggcgca tctcggtca ctgtacccctc caactcccg 48120
gttcaagcga ttctcccttc tttagccccc aagtagccag gattacaggc gcctccatc 48180
atgcctggct aatttttgtt tttttagtag agatggggat tcaccatgtt ggccaggctg 48240
gtctcgactt cctgacttca agtgcgttca cgcgcggatgt ctcccaagt tctgggat 48300
cagggtgtgag ccactgtgcc tggccataat ttttaat(tttt ttttggggat atggaaatctt 48360
gccatattgc ccaggctgg tttgaactcc tgagcttaag tgatccacca gcctcagcc 48420
cccaaagtgc taggactata ggtgtgagcc actgtgtgtg gcctaatagt cacat(tta 48480
atgttcagta tttgtgtgtc tcttagtagag tggatctaga atatccatc cttgcagctt 48540
gtccttctaa atccttctct acacatctat catatatttgc ttttccacc cagattgtt 48600
tatatctatt gttactacac aacttatatt tttaaacgca tttatcgtt tttatattt 48660

```

-continued

tgtatgccttg tctgttgttc acacagttt tgtgtcaact gctgaagatg tcatgattag 48720
caaaaacaga tatgattccc cctctcatgg tccttatcat ctatgtgggc agagtccacc 48780
aaaatatcat gtcagtgaat gtgcaattag acaactgagag aggaatctga aggaagcgt 48840
catggttta tgaacaagga acttgaccta gaccctatgc tgagggaaagg cctctgatct 48900
gaccagaggt cagagggtat gatggctggg ggaacacagt gtttcagca gagcaaata 48960
cctgtgcaaa gacacagctc agatcagagg cggaaggago gtgggtgtgc caggacactg 49020
agtgaagaga caatgtgaag ctggactgca ctgagcatgc gggagggtt atcttcattcc 49080
tcagagcaat ggcaggaaat ttttcatttc cagggaaaga gagcgattta gcactttttt 49140
tttttctttt atctttttt tttttttt ttttttttag acagagtctt actctgtcgc 49200
ccaggctgga gtgcgggtggc acgatctcag ctcactgca cctccgcctc ctgtctccgc 49260
cttccgggta gctgaaatta caattctct gtctcogtct cctgggttagc tggaattag 49320
ggcatgcccc accacacctg gctaattttt gtatTTTAA tagagatggg gtttcacgat 49380
gttggccagg ctggcatttga actcctgacc tcaggtgatc cacctgcctc ggctcccaa 49440
agtgttggga ttgcagggtgt gagccatcgt gcccatttga gatTTTACAC tttttttttt 49500
acaactatac agtggatatg tgccataatt ttttaactc aagacctgtt gatccggttt 49560
catcttgctt ttttaagcaa gcacattgcc ttttttttt ttttttttt ttttttaaag 49620
aaacagtctt gctctgttgc ccacgctgga gggcaatggt atgatcacag ctcgctgtag 49680
cctcaaattt tagatggaat tacaggcaca tgccactgtg cccatagact gctctatgtt 49740
atctccctt tggtttttct tggaaagggg atgaacagag gtcacacatg tgataggaca 49800
gtttgggctg actcaaaagg tttgtgcaca aggcacacat tcagaaaatc acttcgcact 49860
tacctactca ttcatccctt cagcaaacat tcattaaacg gcagcagtgc cccagctgg 49920
ggctgggtct ggggtgcaga ggtgaatcta ccagggtccc agccctcagg tgggtggttc 49980
aaggctctta gggagcgcag aggectgggg tgggtgttcc tggcaccctt cggggccctc 50040
cccatggtga tctttggttt gttctgcctg tgcctctgg tgccactgtc tgacttcagg 50100
gacctgcccgg aatgggagcc caggatccga gaggcttcc gcactgggtga cttggactct 50160
aagcccgacc ccagccggag ctccaggcct taccgagctg aagacaatga ttccatgcc 50220
tctgagggtgg gtcttaggcct gagcaagcat gggattgggt aggaggagag tgactgagcc 50280
ctgtcaaaag cagagtggtc catatatcac ttagctggtg aggatatggg tttggggcat 50340
tccatTTGTT tccaggcttc ttggcagecca agggcagaag gggaaataagg aagtttgggg 50400
cattctcatg gcctgatcca aggaaactta taaactcatc tgcataaatt atttttttttt 50460
ctgtactaaa ttccactggg aatthaacaa ttatgtaat ttaatggaaac tactcatgt 50520
tgttagcagga aggcagcagt ctctatagcc aggcagtttgggttcaatt gtggatccac 50580
tctattctgg ctgtgagatc ttggataagt cattttacct ctctgattct cagtttctc 50640
acatgtaaaa tgaagggtgcc tcacagggtt gttgaaagga ttaaatacat gaaaagccat 50700
agcttaaggc atggccaca ttgggtgagg atgaggagga ggaggaggat gattatgtg 50760
ataatgacac atctctggga ctcagggac ctgagttcaaa gtctcagttc taccagtgtt 50820
cctgtgtgac cttggacaga tttctacattt ctggccacag cttcccccattc tgtaacggga 50880
gggttagagt gaatggtagt tattgaattt cttaggattct ttgatcctca gggattctg 50940

-continued

ggtagttctg agcccttct aagectgctg agcttctct tgcgttcataatgagcctg 51000
aaggcagaaaa aggctctac ggactgggtg gagtatccct gaaaccctga accttgcaga 51060
gtcctatgtat tttccagcc tgagcacctt tctgagcatc tcttggaaaga caactctgtat 51120
gtcccaggaa ggtcagccca tgactatgtg ggaaggagaa gctcaattga agctgatgg 51180
tcagactctc taagttagct tccttggaaac catggaaaaa agagaaaaaa ctgagataaac 51240
ttagttgtca ttttctaaga aaaggaaata ggcctattct ttttttggtt tttttttttt 51300
tagactgagt ctgcgtctgt tgccagctt ggagtgcagt agtgcgtatct tggctcaactg 51360
taacctccat ttccctgagtt caagcgaatc tcgagcctca gccctcccaag ttaggtggaa 51420
ctataggcgt gcgccaccat gcccagctaa ttttctgtatt ttttagtagag acagggtttt 51480
gtcatgtcac ctaggctggc ctcaaactct ggggtcaag tgcatttccc gtctcagecct 51540
tccaaagtgc tggatttaca ggcgtgagcc actgtgccct gccaggaaat aggccatttc 51600
tataggctta cctgttagaaa aaatgatttc ctgcacatc tcaattttt gtttcatca 51660
gccactctt attgagtagt gacaattgcc cagtgcgtt gaccactgag ttcagggagg 51720
aaaccaggaa cagcagacag tatgttagccc tgcaaaaatt ttaatgcaa aacagatata 51780
gaagatatac ctcaagtcac tgcgttcc actgccaat catcatatta ttgaccatat 51840
aatgaatgtt tactttctca cagagggaca tttgccatgc ctaggaaggat tttatgtagg 51900
agttctgggg ccaggcacgg tgactcacac ctgtatatttgc actgactttgg gaggccagg 51960
tggcagatt gcttaagcct aggaggcttgc gatgagcctg ggcaacatgg caaaatccctg 52020
tctctatcaa aaatatacaa aaatttagcca ggcgtgggtgg cgggtgectg taatcccagt 52080
taattgagag gctgaggcag gataatcgct tgaacctggg aggcagaggt tgcagtgc 52140
caagatcacg ctcaaagctg gcgaaactcc gttgaaagag agagagagag agagagagag 52200
agagagagag agggagagag agaaggaggg gaggggtggg gaagggttggg ggtcagaagg 52260
tcttcacata gttccccactg gttacatttta ttgtgtgttgc ccatgttctt tgcattttttt 52320
tttatggagg cctatagcgt ttttggtttt gtttttctac tttctgttgc tgcattttttt 52380
aataacatac atacaataaa ggacactgtc cttaagggttgc ccactgaatg gattgttaca 52440
tatgtataga tccattaagc atcttttaag tcccatcatt tagtatctaa gttgtttctaa 52500
aaatattgca attataaaca atgaaacaat gaacatctttt gtgaatcaaa cttatgtttt 52560
ccagttacattt attaagataa atttcagaga gaaaagaaca agttccaatg gtaacaaaag 52620
ccagatttga tatgtactgtt tcattttaaat catcatgacg accctcagag gtggggacag 52680
tcattctgcc tgatttgtat ttggacagct tgggtccaaat atcccatcgat tcccttacca 52740
ctgtcatctt tagatcacaa ttctatcca ctcttcctc tctctccctc ctcattatctt 52800
gtgtccctgt tctgttaccac agatcaagga gtcgttcgttgc tggctggcttgc aggcccacga 52860
cagcctccgg ggcttgcac agcagcttc ccaggagccg cagctacgaa aggaggaggc 52920
cgacaatttc aaccagaaaa ttggccatgttgc gctgtgggtcc cagcgcttc caggcagcta 52980
ttcctacago ctctctgacg tggaaagcaga ggccttgagg atctatgggc cctctgaatgt 53040
caagaaaatgc aggggaggctt ggggtgggtg gtcacgcctt gtaatcccag cactttggaa 53100
ggccaaggca ggcagatcac ttgaggccag gagtttgaga ccagcctggc caacatggaa 53160
accccgccctc tactaaaaat acaaaaatttta ggcaggcgtt cactagcactt cctgtatcc 53220

-continued

cagctactca ggaggctgag gcaagagaac cgcttgaacc caggaggcgg aggttacagt 53280
gagccgagat catccccactg tattcttagcc tgggtgacag agtgatatact aaaaaaaaaaa 53340
aaaaaaaaaaa agaatgaaag aatgtcaagg gaaacctgag tcttcagccc ctgtgccctg 53400
ctggatgtt attgtccttg ccattgtgca tttagttgtt aacagggctg tcattggcgg 53460
cctgtggac ctggggccag tcctgtttc ttcaggacc tcagttccc atttatgaaa 53520
tggtgccctc atcaggectg gcttctaatac aagataatag ttctggaata acatttgtt 53580
agcccttaact gtatggtggg cctgtccaaat gcactttcta cccattatgt cttaatcc 53640
ttgtgcccctc ccttatcacc atttgcaga tggggaaact gatgcccaga gagctggagt 53700
gctctgcca gggttgcaca tccagagtgt ggcaggcgtg ggcctctaatac ccaggtccct 53760
ctgttaggca tgacactgca ttgcttcata ggggttgcc tagtcatcag ctgtcctcaa 53820
ccctggcacg cccttgctc tccatcgctc accccctccct gctatcatcg ctcaccctcg 53880
caggttcctc attgaacatg agcattcggt tgagacattt ggtgcccagg aggtgctcag 53940
tatatgatgt catggtttt tagagccagg gtcaaatccc agtccttca atgatgtgtc 54000
actgttttg caggagcaga gggcagtgct catctgaaata gattgcgc当地 cgtgtgggt 54060
gatccccggc agggatcaact cacaggcacc acccagactt tcctccctgc ccctgtccag 54120
gggcccagaa aggggtgggg gagaggacag ggcctgggg ttgctcaagc tcttgcctcc 54180
cgccagctga aggaggacca gcagagggcg ctccctgaggg gggagttga gctgcagagt 54240
ctgagcctcc agcggaggct ggagcagaaa ttctggagcc aggagaagaa catgctggtg 54300
caggagtccc agcaattcaa gcacaacttc ctgctgctct tcataaagct cagggtggttc 54360
ctcaagcgct ggccggcaggg caagggtttt cccagcgaag gggatgactt cctcgaggta 54420
agattgggcc agggactggg actggggatgt tgggtgggg agagaggac agggaaaggga 54480
cgccccagctg gtattgacac aactcctagg ctcacatacc tgccatcagc atgcaggctt 54540
ccttctgtca ttctgtgtt tttttttt ttttccgaga tggagtctca ctccagccca 54600
ggcttagagtgt cagtggtgca atctcagect gctgcaatct ccatactcccg ggttcaagca 54660
attctcctgc ctcagcctcc cgagtagctg ggattacagg cacatgccac catgcccagc 54720
taattttgtt attttaagta aagacagggt ttcttggccg ggtgcgggtt ctcacacctg 54780
taatcccagc actttgggag gccgggggg gtggatcagc aggtcaggag atcgagacca 54840
tcctggctaa catggtgaaa ccccgctctc actaaaaata caaaaaatta gccagggttg 54900
gtgggggggtg cctgtgtcc cagetactcg ggagggtgag gcggggagaat ggcgtgaacc 54960
caggaggcgg agcttgcagt gagecttagat tgtgcactg cactccagag tggggagag 55020
agcaagactc cttctcaaaa aaaaaaaaaa aaaaaaaaaa aaaagacagg gtttctccat 55080
gttggccagg ctgatctc aaactcctga cctcaagtaa tccacactgcc ttggcctccc 55140
aaagtgtgtt gattacaggc atgagccacc atgcctggcc aaaagatatg tttacttaac 55200
aagaggaggat aggctggctt aaaaacattt gtgatagttc aggtggattt aggatgtggc 55260
aggaaacgtg gggggggggc tctggaaatg taagaagctt agaaaggacc cacctgaccc 55320
attcatgccc cagtctatgc aggaccctt gctgctcaaa gcccctgcgt cctgtcatca 55380
gcacagcagc acacatcaac acatcaggaa agagaagggg acttacaagg atgtcagacc 55440
ttggtcacac tcatcttagt tgacattgtt cacttggcc gtgtcccttt gggttcagg 55500

-continued

acctcgagtc tagcccaggc ttggccctcc tctagtccag gctgctctct gtagtcccc 55560
gtccagatcc agtaggggagc attggcatgc ctcaagttgc cacttacgcc atcaggggta 55620
ctccagagga gcccagatgc cccaggagcc aggcaagaca gttagacctg agtgtgttagc 55680
aagaccctgc atcaccctat ctcagtgggg ctcatcatte tcgcaggcca ccagctcctc 55740
cccaggccct catggagag tcactaccgc cattccatag gcccacaccc tgctgcccctg 55800
gccctgcagt gcctgacaca ggccgcgcggc tgggagcatt cgccctcggtc ctccctgtcc 55860
aagctgtctc ttctgtgacg tttggaaaca taggacccct ctcacctgtt tggaaagtctc 55920
acaggccctc ggatgaagcc cagtttaagg gaatgacctg caggccacac agtctatatt 55980
ttaagtcagt ttgtctcaga cgaaccctcc acatttggtaaaaaccccttc tagccatttc 56040
catatccttag agtaaaagag agagaaaactt acttttaggaa aaaaaaatta gggcataaac 56100
atagaaaatt ggcttaggtgc agtggctcac acctgttaato tcagcacttt gggaggttaa 56160
ggtgggcaga tcacttgaga tcaggagttc gagaccagtc tggtaacactgatgaaaccc 56220
tgtctctact aaaaatacaa aaattagctg ggaacgggtgg tgcacactgg tagtcccagc 56280
taactcaggag cctgaagcaa gagaatctct tgaacctgca aggcagaggt tgcaactgagc 56340
taagagcatg ccattgcact ctagcctgga tgacacagtg agactccatc tcaaaaaaaaa 56400
aataaaagaga cacagaaaat tacataacat atataattaa gaaatatata catttaccaa 56460
aataaggcata attaataaga tataaacata atcatgctgg gcatggggc tcgcgcctgt 56520
aatcccagca ttttgggagg ctgaggcagg aggattgctt ggcacagaa ggtgactcgc 56580
tccctttgt cttatccca tcctgaacta ggtccctgat ctccctggct ggccctggac 56640
tcttgcccag aatctatcag caagcagtcc aaccccccag caaggtccct tctgtccacc 56700
caaacctttg gtcaccctta tgttagcattt tgggagtgcg tggcccttt tcattcaac 56760
cacagtgcct gccttgaccc agcctgtcaa catggttatt tcagagtgc tctcaggat 56820
gcttctgggt ctttgggca gaatagatgt cgtggaggo taggaggcc tgactatcca 56880
gcccgtctcc tcgacactaca tggtgcctaa gggttggaaatt aaaatgggaaaatcccc 56940
acctgctgtt aagcatggtt tttaaaaaac aggtgaacag catgaaggag ctgtacttg 57000
tgatggagga agaggagata aacgctcagc attctgataa caaggcctgc acgggggaca 57060
gctggaccca gaacacggtg tttccacgc cccttcccg tccctgttg tccattcatt 57120
cctttgtctg ctcagtaatc aattcctact gctttctctg accccattcc tggacattga 57180
agttcaagt tattgcctgg tcatcccagg gcagtaacag agtccagatg atggactgtg 57240
gggtcatagg caagacagca agtaatttcc tttgcagttt aggggacaat atcagggta 57300
aggggtcatg gatgatgggt tttgaaggat gtataggat ttacctaata gagaatagg 57360
gaaagactat tctggaccaa gagaactctg tggccaagg tggcaagg tggctggta 57420
ttgagtatgg cttggaggt acatggtaag ggtaaggata tcaggctaga gtcatactgaa 57480
gagggccctt gtggcttagg gtcaggaacc ttcaaatcaa gccataagtc agagagaaca 57540
gtggggtcag gatcccaggc tcacacacgg tgagctgggg tagggatggg aaatggggc 57600
atggggccag gagagaagat ttgggacaga atcaggcaaa ctgaggagat ctccatgaa 57660
tagacccac atctggcctg cagcgggtca gagaggcaga tcctgagtga ggagaaggac 57720
ccacggcaca gacgctgcag agcctgaag aagccagatg attcagctca gagttaggct 57780

-continued

gaataattat tactgttaact gcagggttt ttcttggtgt cagattggg cctcatgtca 57840
gccttaggagg tagccagggc tggcctgtc attcccattt tatgggttag atgacctagg 57900
gagggttgtgg gacttcgtc aggccctcg aagtcccaag ctggactcct aggtgagttc 57960
tctgaaccca gttccccctc cctacccac agaggctt gactcaggtt ttgcctcctg 58020
tctgcagccc aatgagtaca tcaagacact ggccgacatg aagggtgacgc tgaaggagct 58080
gtgctggctg ctccggatg aacgccgtgg tctgacggag cttcagcaac agtttgccaa 58140
ggccaaaggct acctgggaga cagagccggc agagctcaag ggccataacct cccaggttag 58200
ccccccaccc gtcagacgccc tctccctta ctctcagccca agccctttagc tgtcagatct 58260
gggagtagaa taccaaggcc atgctcttag aggaataaga atgccgttc ttacactgtg 58320
gcccagagag ggcaagagac ttgcctaaag tcacacagca agggatcag aaacgggtgg 58380
gaatgggagt cacagaggaa gggaaaggaaat gtgggggatt gcgcctggg ctcacaggtg 58440
aggctgggtgc acatccccac tgccctgtc cctcatttc tccttcattt cctcataagc 58500
actccctcag ccccaactgc ctggctctat tttgctccct ccacgcaaaa tggggccctc 58560
ccctcccaac tccccagcgt gcccccaaag gagcctaacc caggagccct gggggtttgg 58620
tgtctgtgt gtgcctgagg gtgacttgc gcctccttgc gcctgaaaag tgagagggtg 58680
ggcttcttcc cactcagaga gaatgctctg accctgcago ttggcctcca gtttcttcca 58740
tccagcctgc ccctgtcga agtgccttc ccaggcgtca gcagccctgc ggaggcaggc 58800
actccgatgg tctctgtgg acagatgaca aaactggcac agaggagttt aatttatctc 58860
ccaagggtcac ccggctaata agtgcagac tggcccttgc tcccccaaac ctctaacc 58920
tggctccctg tgaccccaat tccacctct ttttttttttttttttttttggagacgg 58980
catctcaactg tgcactcag gctggagtgc agtggtgcaa tcttggatca ctgcaacctc 59040
cgccctccgg gttcaagaga ttccctgc tcagcctgca aagcaactgg aattatagat 59100
ggcacaccac catgcccagc taattttgtt attttttagta gagatggggc tgccacatgt 59160
tggccaggct agtcttgcac ttctgtatctc aagtgtatctc ccggccctcag ccccaaaag 59220
cgctggggagt actggcgtga gccaccactc ccggccccaa ttccacccctc ttccactgc 59280
tcacagtctc ctggctactt cctggggcca ctactctccc ctcctaacc tcttccagat 59340
ctcacagctc cctccatctg catattccac ggtgtccct gcacccatgtc ccccaaaaga 59400
cctgggcatt tcaggtgctc ctcacccaa gaacctgacc agtcatggg ggttaatggg 59460
gggcattgag tctaaagggtt gcacattccc ttccctgcca ctctgaggtt tgaggccaa 59520
aaaaccatct tgctggaatg accactagta atctttctga gttttatga ggatactgac 59580
atcatgcctg aaaagccac agaagagcat ggcttataaga aagagctctt caaacccctt 59640
ttgtgtggag gggttagggg ttatgtctga gtggggagga ctaagctggg tttagtccctt 59700
gtgaggggtc ttctgagggg acctctcccc ttttccccgc tggcaccctg gagggaaagg 59760
gagagaaggg ccaagaggag tgccctcgtc ctgccagctg cttagtctgc ttcttccgc 59820
agatggagct gaagacaggaa aaggggggccg gggagccggc agggcccgac tggaaggcag 59880
ccctacagcg ggagcgttag gaggcagcgc acctctcgtc tgagtcctac agcgctgtca 59940
tggagctgac tcggcagctg cagatcagtgc agcgcaactg gagccaggaa aagctgcagc 60000
tggtgagcg gctgcagggt gagaaggcgc aggtggagca gcaggtgaag gagctgcaga 60060

-continued

-continued

tgatttgcgg aaggtcataa agtgagtaag ggagaggacc agaatctggc agaactccca 62400
gatctggaaa gttcctggtc ctgggccatg cccactttcc tgcaggtaaa gagtccttt 62460
gagggtcaca cttgcttagga ggctgtgctt ggggagtgcg gatggcagg ctgactaaca 62520
aatcccagaa ccatccaggt accatgtgga aagcaggta gatgcatgca tctgtttgca 62580
ccatcaaacc aaattcctgt ctataaaaga ataattcctt ctttattat taacaacaaa 62640
atgtttctaa aatgattctg aaattatgaa acaagcttca tgcacccctt gggaaacatg 62700
tcagccccaca ggctgtctt gtcagctcat cagggcacgg gttccatgtt ctgtgttgt 62760
tctcacatca gagacttgcc ggcctctag ctccagccctt ttgtgtctct gacattcttc 62820
tgcctcctgc actttaaacc aaatccccca aagcagcatt cttcttaca ctggactgac 62880
gattaagcca gactgggtgaa agcaacagag tggctgggtgg gaaagagctt ctgaccagg 62940
gtccccagtc cttacctggc tccaccaggta caaaccccgat tcaagtcacc tctgtatct 63000
ggccctcatac tgtaaaataa ggtgaccat tcttgacattt ctcattttgc aaggcagcta 63060
tggaaataaa gtgagatagc cagaagattc aaagtgcctt ttaagccac ccctgtctgg 63120
ggcttgggtt gattctatgg tcttcatctc tcattgtgtt ttgtgtatgt tagtgttatt 63180
tccctcccttc cttccctccctt cttccctccctt cccttcttctt ttcttcgtttt ctttcttctt 63240
ttattttagt acaaggcttt gctctgtcac ctggctgga gtgcagtgcc accatctcag 63300
ctcaactgca cctccgcctc ccaggctcaatgtatcttcc cacctcaacc tctctgtcag 63360
ctgggagcac actctaccat gcctggctaa cttttatata tttttttttt tttttttttt 63420
ttttttttt tttttttttt atggatgttc actctttgtt cccagacttag agtgcataatgg 63480
cgcgatctt gctcaccgcac acctctgcctt cccgcgttca aacgattctc ctgcctcagc 63540
ctcctgagca gctgggatata caggcttata ccaccacgccc tgtctaattt tttgtatattt 63600
tagtagagac agggtttctc catgttggtc aggctggctt cgaactccca acctcagggt 63660
atccacactgc ctggccctcc caaaatgtctg ggattacagg tggatggccac cgtgcccggc 63720
aactttgttata tttttttttt gagatgggggtt ttcggcggtt tgcccggtt gatctttaaac 63780
tgctaggctc aagtgtatctt cccacactgc cctccaaag tgcgtggattt atagttgtg 63840
gctgccacac ctggccctgtt gttatatttca ttgttgcctt tattactgtt tgcgtatgt 63900
tgagtttctt acaagcaag cccttagacat gaattcaagt acaagtaagg tgaccaaccc 63960
atgctggctt gcctggactt gagegcgttc ctggactttt gactttttaa actagaatag 64020
tccttagccgtt agcctacttag tttttttttt aacttgagggt tctgggacat gggactttca 64080
atgcaaaacac tgagaaatgc ctggccaaac caagatgagt tggttaccctt gtgtgcaagt 64140
gattttttttttaa cggatggagt caggatgggg aaggggaaga aaccacgtaa gatatgtttt 64200
cagatgaagt ctggccctgg cctgacccca tgcacatgccc aggaggtaa acagtaaatc 64260
accccacagt ttgggtccacc ttaaggcaag ggggtctgcctt tttggggccgg tctttggac 64320
ccctgtctttt tgcgtatgtt ggcgttgattt cggttctaat tgcggccaggc aatccctccag 64380
agaaggctgtt agtgcagat ggtacagat gagccaccaaa caatgggaag tgctacacga 64440
tcatttactc agtgcagatgc ccaacactttt ctctcgtctt tgatccctctc atctgcaaa 64500
tggaaataac cacaagaaca cctatcatac actgtttctg tgaggattaa atcaactaat 64560
gccccccagac gccggcccttc acacactgtaa tccacccat ttaggaagtc gaggcaggcg 64620

-continued

gatcacctga gtcaggagtt caagaccagg ctggccaaca tgtgaaaccc gtctctacta 64680
aaaacacaaa aattagccag gcgagggtggc tcatccctgt aaatgccagc cactcgggg 64740
gctgaggcag gagaatttgt tgaacctggg aggcagagtt tgcaagtggc tgcgatcgta 64800
ccacagcact ccagtcttga caaacagatga agaccctgtc taaaaaaaaaaaaattta 64860
actaatgtgt gcataggaga taggttccat cacaagccac ctggaaagtga tggagctcca 64920
aagaggggag agaagcagag gccctcaag gcccagcagg aaaaagtgcg gagatccata 64980
aatgacggcc actgtggatg gaggcgtggg ggtgccttga tcttacattt gccactctg 65040
tgggatgttag ctgctgtta gctagtcttag ctgttctgtg aattgtaaaca tgaattcaca 65100
tagtcatggc tggtaactata tttaggtataa tgacatcttag aaggatgcct gcaagtagtt 65160
ccttttatta tcattttcca aatccaccag gcagagtggg ggcgtctgag ctgttacccca 65220
ctgagaaggt ggccccctgtt ccatgtatgtt ggttctgtgg gctgcagagt ggagtctggg 65280
gaatggccctt gcttagagca gggccttgc taccagcttt gcccattacc gaccacccct 65340
ctctgcccctt gcccagctgc agaaggctgc cgaccctgg gtcctgaagc actcggagct 65400
ggagaagcag gacaacagct ggaaggaggt gagtgccggc gctgctgcca tgctatctt 65460
gctcctggct gcaccataaa ggtccctgt tgggggggac atgagaggaa gttgctaacc 65520
ctgtctcata ggtcacacccc ttgcacccctt ggctcctcag ccctggagtc tggatctcat 65580
ggttctgtcc tcaaaaaaaaaaaaa cccaaaccaco cagggactgg gctaaagttag 65640
ggggtaggggt tgaggaggaa gaagacaagc ccaccaccag ctgagcgttg gaccactgg 65700
tgctttgttg cagtcaccac atttggatta aattttaaat ggtccctttc cttgggtggg 65760
tcactcttgg agctctgatt ggcccagct ggatcacata tgggggggac gttggatgca 65820
gggcagccag aaactgacca tgcttagatca ttgatgacat cattcctgtc ctccggatga 65880
ggagattcaa gcaactgaggt cggggctggt ctcttgcattt atgccttgc tggctctgg 65940
gccagaactg acccttctg tctctttcc ttgacagaca cgcgttgaga agatccacga 66000
caaggaggct gttccgaag ttgagcttgg aggaaatggt ttaaagaggat gcttgcattc 66060
tggccccctg ctctgcctct ctgaccggcc ccacagcctc gccccttcc acctcctgtc 66120
gcccctgca cccactctac ctgtttcac cgcagccctg ccccgccccctt tggcccttt 66180
cctggcccca ccccaccgca tccaaatggg gagagcatcg ggatgggtt tctctggaaac 66240
ccgagccact caggagctgg aaggcttggg gcagtgggaa ttccaggctg ctgcatgttg 66300
tttggatccg ggaatggaaac acagggtggaa acaaaacatt ttgtctcttct tgggattctt 66360
ttctctgtcc tctggagttt ggagagtcgatttgcatttgcatttggg attattggac 66420
agcatccctt gggcccgcc ctgtgttatgc gccaaggaga ccccccccccc aagaccataa 66480
ggaccacgctc tccactgtta gggaaacccat gaaaggagtc agggaaacag aaccccccac 66540
cgctatcaga ccagcccgaa agcaaggcag gcggggacagg ctgtggagcc agagggccag 66600
gtttgaatgc cagccctgccc agttcttagct ctatgacattt gacccgtggc agtacttccc 66660
atctctgtct agtggggatg ataggagctg cctcttgcatttgcatttggg attattggac 66720
cgtgcatttc catatgcctg gtgtgttcag tgattcaggc aatagggtgtt ccaggagtt 66780
catggcgctt tagttgggt tccccataa gcaaaaccctg agacaaagat ttaagtgc 66840
atggtttattt tgggacagaa ttccaggaaa cacttggaaag tgagacaggg aagggaaagc 66900

-continued

agccgataaa aggagcttca accaggttat tccacggca actggagttc aggtgtcgg 66960
aggaactcag ggagccagag gagaacatgc agcttagagt catcccacca caaggaggct 67020
ggggcacattt ttcaccatatt cccacattt gtttgttag ggtgtgttt aggggtgtga 67080
actcattaac tctgtgttat ctctagttc ggttggcg tgtgtttgt ttttttgtt 67140
tttgtttttttt ttttgggca gagtttgct ctgtctcca gggtgaggta caatggcaca 67200
atcttagttt actgcaacct ccacccca ggtcaagca attctctgc ctcagccctc 67260
tgagtagctg gggttacagg ctaacgccac cacgttggc taattttgtt attttacta 67320
agagataggg ttttaccatc ttggtcaggc tagtctcaaa ctccgtaccc tgggtgatct 67380
gcctgcctca gcattccaaa gtgtgggtc tagttcttc tatgtgttagg acagccaagt 67440
atgatccac aggcagaaaa aaaaaccatc agatagagaa tggtgaggta tgcatgtcag 67500
agtatgcagt ttttaggtgta taggtgaaag tgcaaagaga ggccaggtgc agtggctcat 67560
gcctctaatac ccagcatttt gagaggcaga ggcaggaaga tctcttgagg ccagaagttt 67620
gagaccagcc tggcaacat agggagatcc ccatctctac aaaaattttt ttaaaaattta 67680
gcctgggtt aggctggca cagtggctca catctgtaat cccagcactt tgggaggctg 67740
aggcgggtgg atcacctgag gtcaggagtt ccagaccgc ctggccaaca tggtgaaacc 67800
ccgtctctac taaaatacata aaaatttagct gggcgtggc ggcacgcct gtaatttcat 67860
ctgctccgga tgctgagaca ggagaatctc ttgaacctgg agggtggaaatg ttgtgttag 67920
ccgagatcat gccactgcac tccagcctgg gtgacagaac aagactctgt ctcaagaaaa 67980
aaaaaaagaaa aaaaaaaagcc aggttgggtg gtgcacaccc ttagtcccaag ctactctgg 68040
ggctgaaatg ggaggatcac ttgagccag gagtttgagg ctgcagttagt ctattattgc 68100
actactgcat tccagcctgg gtgacatatac aagacctgtt tgggagaaaa aaaaaaaaaaa 68160
gaaaatgcag acagaatgtg ggactggca ccaacaacat ctctacaaga ggtgaacaag 68220
acggtcctgg tctttgcct cacgcageac acggaccagg gtggtagacc agagtgtgcc 68280
ctcagtgtta tttccactaa taacgacgtt tccctctgt tccttcgcag aaccaaatct 68340
gtttttcca tgtctgagtt tgaaagttt ctcgactgtt ccccttacct tgctggcgg 68400
gatgccccgg gcaagaagct gcctaacaac cctgcctttg gcttggtagt ctccgagcca 68460
ggggatccag agaaagacac caaggagaag cctggctct cgtcgaggga ctgcacccac 68520
ctgggtgccc tggcctgcca ggaccccca gggaggcaga tgcagcgcag ctacacggct 68580
cctgacaaga cgggcattcg agtctactat agtcccccg tggcccgccg cctcggagtc 68640
cctgtggttc atgacaaaga gggcaagatc attatcgac ccggcttcct cttcaccaca 68700
gccaagccca aagagtccgc cgaggctgtat gggctggctg agagcttca tggtcggtag 68760
ctctgcaact tctcacggca ggcgcctggac ggaggcttag cgggcagccc ctcggcggcc 68820
gggcctggct tcccagccgc cctgcgtac tttgagatgt caggcaacat gagtgtgac 68880
atgaaggaga tcaccaactg tggcgccag gccatgcgt cggctactt ggagaggaaa 68940
gtgaagagca catccagcca gacgggtggc ctggccagtg tgggcacaca gaccatccgc 69000
acggtcagcg tggcctcgca gaccgaccca ccccgacgc gcctccatgg caaggcctgg 69060
tcaccccgca gctcttcgtc cgtgtctgtc cgcagcaagc agatctctc ctcctggac 69120
aaggccatt cgccatcgca gccggccctgc tgctccccc agtatggctc accaaagctc 69180

-continued

cagaggcggt ctgtgtccaa gctggacagc agcaaggacc gcagcctgtg gaacctgcac 69240
cagggcaagc agaacggcgc ggcctggcc cgctccacca ccacgcggga cagccctgtta 69300
ttgagaaaaca tcaacgatgg actctccagc ctcttcagtg tggtggagca ctcagggagc 69360
acggagtctg tctggaaact aggcatgtct gagacgcggg ccaagcccga gcctcccaag 69420
tacggcattt tgccaggaatt ctccgtaat gtgtgtggcc gggcacccgag ccccacctca 69480
tcagcaggag aggagggcac caagaagcca gagccccctt ccccagccag ctaccatca 69540
ccagagggtg tggccaggat cctgaacaag aaggcagcca agttggcagc cagtggaggag 69600
gtcagactca ccatgctccc ccagggtgggg aaggatggtg tcctccggga cggagatgg 69660
gccgtggtcc ttcccaatga ggttaggtggg tgggatctgt ctttctctt agtaggtgg 69720
gttagtatata aaggtccaag ctcttttgtt tttaagttc agagatatgg gccgagggtgg 69780
gtggatcgcc tgagatcagg agttcaagac cagtctggcc aacatggta aaccccatct 69840
ctactaaaaa tacaaaaatt ggccaggcgt ggtggcatgc acctgttgc ccagctgctc 69900
tggaggctga ggttaggagaa tcacttgaac ccagtaggcg gaagttgcag tgagctgaaa 69960
ttgtgttattt gcactccgcg ctgggtgaca gactaagact ccgtctcacca aaaaaaaaaa 70020
aattatctat ctgtctgtct gtctgtctgt ctgtctgtct gtctgtctgt ctatctatct 70080
atctatctat ctatatatct atatatataa aataaaaatt cagggatatg gccatgttt 70140
ccaggaagtt ctatgttgg tctaggatag agtttggaaagg tcccagacct ggaacaacac 70200
agacctgatg gtcacatggt ctgaattcca agagactgt gacatggct taactagcag 70260
ctttgtctcc acccaccaga gactgccagc atccccctca gggaaaggggc cttactgctc 70320
tcctgccacc cagtggtgtg cttaggccag cttgtttggc ttgcaggagc caattgtcag 70380
catctctgtg ggtttttttt tttttattgt ttttggggg gtttttagag aagtctcggt 70440
cttggccccc aggcttgagt gcaatgactc aatctcggtc cactgttaacc tccgcctct 70500
gggttcaaacc aattctgtct ctgectccca agtagtggg attaaggcac ctgcacccat 70560
gccccggctaa tttttgtatt ttttagttaga gacggggttt caccatgttg gccaggctgg 70620
tctcgaactc ccgaccta gtatccgccc caccttggcc tcccaaagtg ttaggattac 70680
aggcgtgagc caccggccct ggccacatct ctgtgtttaa catcaccttg gtaccttgag 70740
attggccaca gtggggaggat ttacagcata aaaattggta aacactacaa atcggggctc 70800
ctcctatcac tcctggaaaa ctaattgtta aacattagcg ggcataaccac ttctccaccc 70860
catcaagact tcacaccta gtgtcacat ttccctgttag aatctttcac accccactcc 70920
tgctatcagt ttcttataag ttaatttagta aaccttattt gaatgttagt gacatttt 70980
tgaatgtgtt agaatgtgca taagttact ttgcagaaag ggaaggaaat gtatttgaag 71040
gatggcatgt ttataaaac ccaagagtagt gacccatgtaca acctcagcag ctttcctctg 71100
tcttcctctc ctctctttaa gttcacccct ctctgtatatac atattctcc tttttctctg 71160
tttctctctc tgggtgtttt gcaacaggac tgggtgggct gtcccgatctc tgactttaca 71220
acatcttcca gttcaagccc ctaggagaga ctgacacaaa tccctgtctc ccaattccac 71280
atttctgaaa gaaggaaact ctgatgggtg cagctagagg cgagtgtacc ctttggcc 71340
gtctatggct gggagtggtt ccatgtggcc tgctgtgggg cacagagaca tgagaaggcc 71400
tgtgtgtggaa atggggctgtg atagactcaa acattctact tttcttattt tgtacgttt 71460

-continued

gggtggttct agtttgcaccaat aatgctgtgg tgaacatctt ggtgc当地
catttttta catttcagac tattttcccta ggtataagtt gcagaaatga gattaaagga 71580
taatagaata taaaatgtggc cgggcgcggg ggctcatgcc tgtaatccca gcaactttggg 71640
aggcccgaggc ggggtggatca caaggtcagg agtttgagat cagccctggcc aatatggtga 71700
aaccctgtct gtactaaaaaa tacaaaaatt agccgggcat agtggtggtc acctgttagtc 71760
ctagctactc aggaggctga ggcaggagaa tcgcgtgaac ccaggaggcg gaggctgcag 71820
tgagctgaga tggtaccact gcactccagc ctgggtgaca gagtgagact ccatctccac 71880
aaaaaaaaaa aaatataaat gttttaagt tatttgataa atactgc当地 attgttttc 71940
aaaaggtgtgc catatctaca gcttccagca gttgtggcg agtgc当地 ttcttctata 72000
tctcaactcag ctggctgaca tttaaaaaca tcttttaaat atttattcct ttttggcca 72060
ggtgcaatgg ctcatgtctc taatccc当地 accttgggag actgaggctgg gcagatcact 72120
tgagcttagga gttcaagacc agtctgagca acatggtgag agtctcatctc tacattaaaa 72180
aaaaaaaaaa agaggccagg catggtggt tacggctgta atcctagcac tttgggaggc 72240
cgaggcagggt ggattgcctg agtc当地 agtggccaa gcctggccaa cacagtgaaa 72300
ccccatctct actaaaaaaat acaaaaaatt agccaggcgt ggc当地 catgc gctatagtc 72360
ccagctactc gggaggctga ggcaggagaa ttgcttgaag cc当地 gggagggt gaggcttgc当地 72420
caagctgaga ccacgccc当地 ctactctactc ctgggcaaca gagcaagact ccatcttta 72480
aaaaaaaaaa agggccc当地 agggccaggc gtggtggtc acacctgtaa tcccagact 72540
ttgggaggcc gagagaggtg gattgttga agccaggagt ttgagaccag cctggccaa 72600
gtagcaaaac cctgtctctc cttaaaaat acaaaaaattt gcccggccca gttggctc当地 72660
cctgtatcc cagcacttt ggaggctgag gtgggtggat catgaggctca ggagatcag 72720
accatctgg ctaacacagt gaaaccctgt ctctactaaa aataaaaaa gttageccagg 72780
cgtggtgcc ggc当地 cctgta gtccctagca ctgggaaagc tgaggccagg gaatggcatg 72840
aaccggggag gc当地 ggagcttg cagtgagccg agatc当地 gccc actc当地 agcctgggtg 72900
acagagcagg actccatctc aaaaaaaaaa aaaaaaaaaa atacaaaaat tagctggca 72960
tgggtggcaca tgcttacagt tccagccact tgggtgatgg aggcatgaga attgtggaa 73020
cccaggaggc agtgc当地 gca tatcgc当地 ctgcactcc gc当地 ctgaggccaa cagagtgaga 73080
ctctgtctca aaaatataata aataataaaag tctcaaggca gggaaaagc aagtgtctca 73140
gtctgaaggc tgc当地 caggcag gaagaattct cttacttgag ggagagtaag cc当地 ctgtt 73200
ctgttcaagg cttaactga ctggatgggg tccatgttag ggagaattcg cttttttttt 73260
tttttggagac ggagcctcactc tctgtcaccctt aggttggaaat gc当地 gtgggtg gatctcg 73320
caactgc当地 accctcactcc aggctcaaggc gattcttgc当地 cctcacttcc ccaaccaact 73380
agctgggattt acaggccat gccaccaccc cccggctactt tttttttttt agtagaaata 73440
gggtttcacc atgttggccca ggctggctc当地 gaactgctaa tctcaaggtaa tccgc当地 ctg 73500
tc当地 agtctccc aaagtggtg gataacagac gtgagccact ggc当地 ctggctg gagaattac 73560
atgttaatct catctcaagg cactctgaca gaaacaccctt gaaacatgtt ttacgtaata 73620
tctcgccacc tgc当地 gggccagg tc当地 atttgaa tataaaaat accatcactt ccatataacc 73680
agcaccacat caaqaaacaa cagcacaacaa ctggctggat tttqaatgtt gattttaag 73740

-continued

ttgatccaaa tattgttcag tatgtaaaat gtgcgtgcta cccattgacc cagtagttct 73800
tctttacga aggttagccta tggaaatatt cacagagatg gataaagaaa aataaaaaga 73860
ggctgggcgc agtggctcac ccaatcacct gaggtcgaaa gttttagacc agcctgacca 73920
atgccgagaa accccacatc tactaaaaat acacaattag cggggcgtgg tggcacatgc 73980
ctgttaatccc agotactcgg gaggctgagg caggagaato gctcgaaccc gggaggcaga 74040
ggttgctgtg agctgagatc gcatttcgc actccagct gggcaacaag agctaaactc 74100
tgtctcaaaa aaaaaaaaaa aaagaaaaga aaagaaaaag aaaaaaagag agaggggagg 74160
aaatcaaagc tcaagagagg ttaagtaat ctccccaggt cacacagctt gtaattggca 74220
aagctggat ttaaccaagc agtttgctc tatcacagca cttttatattt atttatttat 74280
ttatattat ttgagacgga gtctcgctct gttgcctagg ctggagtgcgtca gtggcacat 74340
cttggctcac tgcaagctcc gcctccagg ttcacgccc tctctgcct cagccctcta 74400
agtagctggg actacaggca cctgccaccc cgccggctt atttttgtt ttttttagtag 74460
agacgggggtt tcagatggtc tcgatctct gacctcgtt gtcacccgccc ttggccctcc 74520
aaagtgcgtgg gattacaggc atgagccacc gggccggggc cattttattt attttttaaa 74580
gagatgggggtt ctgcgtctat tgctcaagcc agagcccagt gacacaatca tagctcactg 74640
ctgccttgac ctctgggtt caaaggatcc tccctgtctca gcctcccaca tggccagcc 74700
cagcagcaact tttttttttt tttttttttt agatggagtc tcactctgtc cccaggctg 74760
gagtcagtg atgtgatctc ggctcaactgc aacctccgccc tcctgagttt aagecgttct 74820
gctgcttcgg tctcccaagt agctggact gcaggaatgt gccaccatgc ccggtaatt 74880
tttgtatattt gagtagagat ggggtttcac catgtggcc accctgggtct tgaactcctg 74940
acctcaggtt atccacactgc catagccctcc caaagtgcgtt gcattacagg tgtgagccac 75000
catggccagc tgcgtccagcact tttttttttt tttttttttt ttttcagatg gagtcttgc 75060
ctttcaccca ggctggagtg cagttggcaca atctcagctc actgcaacctt ccgcacccctg 75120
ggtttaagcg attctcctgc cctagccctcc cgagtagctg ggactacagg cgcatgccac 75180
catgcccagc taatttttgtt attttttagta gagatgggggtt ttcaccatgt tggccaggat 75240
ggctcgtatc tcctgaccc tgcgtccacc tgcctcagcc tcccaaagtgc ttaggattac 75300
aggcatgagc caccgtgcctt ggcagcagca cttttgtca ttattttgtt ttattggtaa 75360
tagtaaaaaaa ttagaaacta cccagatgtt cataaaaagt gaattggggc caggcacatg 75420
ggctcgtatc tgcaatccca ccactttggaa aggccgagggc gggcagatca cttagggtca 75480
ggagttttagt cctgaccaac atgtaaaat cccctctcta ctaaaaatac aaaaatttgc 75540
taggcattgtt ggcacacgct tgtaatcttta gtcacttggg aggctggggc acaagaatca 75600
cttgaatccg tgaggtggag tttgtgttgcgtca gtcacatgc taccactgc ctccagcctg 75660
gggaacagag caagattctg tctaaaaat aaaaaataa aaaaagcgaa ttgggtatata 75720
aaatcagtttta tttctatctca tgaaatggt acagagacac taaaaaagaa ggctgatcgc 75780
caggecatggt ggctcatgcc tttggggagc atgagcaccc aaaggtgctt aatccagcag 75840
tttggggagggc tgaggcagga ggatcgttgc agcccaggag ttcaagaccatc gcctgagca 75900
catagtgaga ccctgtctca aacgaaaaaa aaggataaaa aggaaggaga tcttagacata 75960
ccttagtgaca gaaaagtgtt cgtgttttac agttgcaaaa taagctgtca ttttggtaaa 76020

-continued

```
tgtatgttaa gaatctttt ggctgcagg accatggaca gtgtccactg tcctctctgt 76080
atagcctaga atatattaca atgaacttcc attaccgctc taatggaga agaacaacag 76140
aaagaaaaaa ttggtccagt gaggagctc agccctggg aactgggta tggccgagtc 76200
tcagttcctt gcctggata ccacacagct catggtggtc catgtctcctt ggctcttt 76260
gacttctgc ctcctgcct ccctttctt aggacgctgt ttgtactgt agtacccagt 76320
ctctcacctc ctgcttcgccc cgatcgcccc gctctgccat ccggcaactct cttccaagt 76380
gcaggctgca cccttcagag tccagctggg gtggggagga gagggcactc ccccccagcg 76440
agtacagag cagccaaagct ccccgccctca accagccccag cccctggata gcagaaggaa 76500
accagcagag acgagacgag gtgaggcgag gggctgtgtc ctcagcattt cttggccctg 76560
gagggacagc agtgatgcca ctggcagaat gcagcttca catcaaggtt aagccgggtc 76620
tcctgtggc ccctgggtgg tgagcttca cttcccaggg gaaggcagtg agtgggagag 76680
agaccaaacc tgggcttccc aagcatccac tgagagatct gtcaagagcc gatccctggg 76740
tcctaagaga gacccctggcc tggttctgcc c 76771
```

1. An isolated polynucleotide selected from the group consisting of:

- (a) a polynucleotide comprising a nucleotide sequence at least 70% identical to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3 and encoding a functional Suppressor of Glucose by Autophagy (SOGA) polypeptide;
- (b) a polynucleotide that hybridizes to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1 and 3 under stringent hybridization conditions and encodes a functional SOGA polypeptide;
- (c) a polynucleotide encoding a functional SOGA polypeptide comprising an amino acid sequence at least 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 4; and
- (d) a functional fragment of any of (a) to (c).

2. The isolated polynucleotide of claim 1, wherein said isolated polynucleotide is selected from the group consisting of:

- (a) a polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS: 1 and 3 or a fragment thereof that encodes a functional SOGA polypeptide;
- (b) a polynucleotide encoding a functional SOGA polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 4 or a functional fragment thereof; and
- (c) a polynucleotide comprising a nucleotide sequence that differs from the nucleotide sequences of (a) or (b) above due to the degeneracy of the genetic code.

3-5. (canceled)

6. A vector comprising the isolated polynucleotide of claim

1.

7. A cell comprising the isolated polynucleotide of claim **1**.

8. (canceled)

9. An isolated polypeptide or functional fragment thereof encoded by the isolated polynucleotide of claim **1**.

10. A cell comprising the isolated polypeptide or functional fragment thereof of claim **9**.

11. A fusion protein comprising the polypeptide or functional fragment thereof of claim **9**.

12-13. (canceled)

14. An antibody or antibody fragment that specifically binds to the polypeptide of claim **9**.

15. An antisense oligonucleotide, ribozyme, or siRNA that specifically binds to the polynucleotide of claim **1**.

16. A pharmaceutical composition comprising the polynucleotide of claim **1** and a pharmaceutically acceptable carrier.

17. A pharmaceutical composition comprising the polypeptide of claim **9** and a pharmaceutically acceptable carrier.

18. A pharmaceutical composition comprising the antibody of claim **14** and a pharmaceutically acceptable carrier.

19. A pharmaceutical composition comprising the anti-sense oligonucleotide, ribozyme, or siRNA of claim **15** and a pharmaceutically acceptable carrier.

20. A non-human animal genetically modified to express the polypeptide of claim **9**.

21. A non-human animal genetically modified to inhibit expression of a SOGA polypeptide.

22. A method of decreasing glucose production or autophagy in a cell, comprising contacting said cell with the polynucleotide of claim **1** in an amount effective to decrease glucose production or autophagy in said cell.

23. A method of decreasing glucose production or autophagy in a cell, comprising contacting said cell with the polypeptide of claim **9** in an amount effective to decrease glucose production or autophagy in said cell.

24. A method of decreasing blood glucose levels or increasing insulin sensitivity in a subject, comprising delivering to said subject the polynucleotide of claim **1** in an amount effective to decrease the blood glucose levels or increase insulin sensitivity in said subject.

25. A method of decreasing blood glucose levels or increasing insulin sensitivity in a subject, comprising delivering to said subject the polypeptide of claim 9 in an amount effective to increase insulin sensitivity in said subject.

26. A method of increasing glucose production or autophagy in a cell, comprising contacting said cell with an agent that decreases the activity of the polypeptide of claim 9 in an amount effective to increase glucose production or autophagy in said cell.

27. (canceled)

28. A method of increasing blood glucose levels or decreasing insulin sensitivity in a subject, comprising delivering to said subject an agent that decreases the activity of the polypeptide of claim 9 in an amount effective to increase the blood glucose levels or decrease insulin sensitivity in said subject.

29-31. (canceled)

32. A method of measuring the response of a subject to a treatment for diabetes or predicting the clinical outcome of a diabetes treatment in a subject, comprising determining the circulating level of a SOGA polypeptide or a functional fragment thereof in said subject after administration of the treatment and comparing it to the circulating level of the polypeptide or a functional fragment thereof of claim 9 in said subject before administration of the treatment.

33. (canceled)

34. A method of identifying an agent that binds to or modulates the activity of a SOGA polypeptide or a functional fragment thereof, comprising:

contacting the polypeptide or a functional fragment thereof with a test agent under conditions whereby binding between the polypeptide or a functional fragment thereof and the test agent can occur or modulation of the activity of the polypeptide or a functional fragment thereof can occur; and

detecting binding between the polypeptide or a functional fragment thereof and the test agent or detecting modulation of the activity of the polypeptide or a functional fragment thereof upon contact with the test agent as compared to activity of the polypeptide or a functional fragment thereof in the absence of contact with the test agent.

35-41. (canceled)

42. A kit comprising a reagent for determining the expression and/or activity of the polynucleotide of claim 1.

43. A kit comprising a reagent for determining the expression and/or activity of the polypeptide of claim 9.

* * * * *