



US 20060074018A1

(19) **United States**

(12) **Patent Application Publication**
Bernard et al.

(10) **Pub. No.: US 2006/0074018 A1**

(43) **Pub. Date: Apr. 6, 2006**

(54) **METHODS OF DIAGNOSING & TREATING
DIABETES AND INSULIN RESISTANCE**

(75) Inventors: **Allan Bernard**, San Francisco, CA
(US); **Francine Gregoire**, Lafayette,
CA (US); **Brian Lavan**, San Francisco,
CA (US); **Shonna Moodie**, San
Francisco, CA (US)

Correspondence Address:
Jean M Lockyer
Patent Agent
Townsend and Townsend and Crew
Two Embarcadero Center 8th Floor
San Francisco, CA 94111 (US)

(73) Assignee: **Metabolex Inc**, Hayward, CA

(21) Appl. No.: **10/516,635**

(22) PCT Filed: **Jun. 4, 2003**

(86) PCT No.: **PCT/US03/17725**

Related U.S. Application Data

(60) Provisional application No. 60/386,085, filed on Jun.
4, 2002. Provisional application No. 60/386,331, filed
on Jun. 5, 2002.

Publication Classification

(51) **Int. Cl.**
A61K 38/22 (2006.01)
G01N 33/53 (2006.01)
(52) **U.S. Cl.** **514/12; 435/7.1**

(57) **ABSTRACT**

The present invention provides compositions and methods
for diagnosing and treating diabetes and insulin resistance.
In particular, the invention provides methods of identifying
modulators of the polynucleotides or polypeptides of the
invention and using those modulators to treat diabetes, as
well as methods of diagnosing diabetes by measuring the
levels of the polynucleotides or polypeptides of the inven-
tion in a patient.

METHODS OF DIAGNOSING & TREATING DIABETES AND INSULIN RESISTANCE

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Patent Application No. 60/386,085, filed Jun. 4, 2002 and U.S. Provisional Patent Application No. 60/386,331, Jun. 5, 2002, each of which are incorporated by reference.

BACKGROUND OF THE INVENTION

[0002] Diabetes mellitus can be divided into two clinical syndromes, Type 1 and Type 2 diabetes mellitus. Type 1, or insulin-dependent diabetes mellitus (IDDM), is a chronic autoimmune disease characterized by the extensive loss of beta cells in the pancreatic Islets of Langerhans, which produce insulin. As these cells are progressively destroyed, the amount of secreted insulin decreases, eventually leading to hyperglycemia (abnormally high level of glucose in the blood) when the amount of secreted insulin drops below the level required for euglycemia (normal blood glucose level). Although the exact trigger for this immune response is not known, patients with IDDM have high levels of antibodies against proteins expressed in pancreatic beta cells. However, not all patients with high levels of these antibodies develop IDDM.

[0003] Type 2 diabetes (also referred to as non-insulin dependent diabetes mellitus (NIDDM) develops when muscle, fat and liver cells fail to respond normally to insulin. This failure to respond (called insulin resistance) may be due to reduced numbers of insulin receptors on these cells, or a dysfunction of signaling pathways within the cells, or both. The beta cells initially compensate for this insulin resistance by increasing insulin output. Over time, these cells become unable to produce enough insulin to maintain normal glucose levels, indicating progression to Type 2 diabetes.

[0004] Type 2 diabetes is brought on by a combination of genetic and acquired risk factors—including a high-fat diet, lack of exercise, and aging. Worldwide, Type 2 diabetes has become an epidemic, driven by increases in obesity and a sedentary lifestyle, widespread adoption of western dietary habits, and the general aging of the population in many countries. In 1985, an estimated 30 million people worldwide had diabetes—by 2000, this figure had increased 5-fold, to an estimated 154 million people. The number of people with diabetes is expected to double between now and 2025, to about 300 million.

[0005] Type 2 diabetes is a complex disease characterized by defects in glucose and lipid metabolism. Typically there are perturbations in many metabolic parameters including increases in fasting plasma glucose levels, free fatty acid levels and triglyceride levels, as well as a decrease in the ratio of HDL/LDL. As discussed above, one of the principal underlying causes of diabetes is thought to be an increase in insulin resistance in peripheral tissues, principally muscle and fat.

[0006] Therapies aimed at reducing peripheral insulin resistance are available. The most relevant to this invention are drugs of the thiazolidinedione (TZD) class namely troglitazone, pioglitazone, and rosiglitazone. In the US these

have been marketed under the names Rezulin™, Avandia™ and Actos™, respectively. The principal effect of these drugs is to improve glucose homeostasis. Notably in diabetics treated with TZDs there are increases in peripheral glucose disposal rates indicative of increased insulin sensitivity in both muscle and fat.

[0007] The molecular target of TZDs is a member of the PPAR family of ligand-activated transcription factors called PPAR gamma. This transcription factor is highly expressed in adipose tissue with much lower levels being observed in muscle. Binding of TZDs to PPAR gamma in target cells and tissues such as fat and muscle brings about a change in gene expression. The link between TZD-altered gene expression in fat and muscle and increased insulin sensitivity is unknown. The present invention addresses this and other problems.

BRIEF SUMMARY OF THE INVENTION

[0008] The present invention provides methods for identifying an agent for treating a diabetic or pre-diabetic individual. In some embodiments, the methods comprise the steps of: (i) contacting an agent to a mixture comprising a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14; and (ii) selecting an agent that modulates the expression or activity of the polypeptide or that binds to the polypeptide, thereby identifying an agent for treating a diabetic or pre-diabetic individual. In some embodiments, the methods further comprise selecting an agent that modulates insulin sensitivity.

[0009] In one embodiment, step (ii) comprises selecting an agent that modulates expression of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that modulates the activity of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that specifically binds to the polypeptide. In some embodiments, the polypeptide is expressed in a cell and the cell is contacted with the agent.

[0010] The present invention also provides methods of treating a diabetic or pre-diabetic animal. In some embodiments, the methods comprise administering to the animal a therapeutically effective amount of an agent identified as described above. In some embodiments, the agent is an antibody. In some embodiments, the antibody is a monoclonal antibody. In some embodiments, the animal is a human.

[0011] The present invention also provides methods of introducing an expression cassette into a cell. In some embodiments, the methods comprise introducing into the cell an expression cassette comprising a promoter operably linked to a polynucleotide encoding a polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14.

[0012] In some embodiments, the cell is selected from the group consisting of an adipocyte and a skeletal muscle cell.

[0013] In some embodiments, the methods further comprising introducing the cell into a human. In some embodi-

ments, the human is diabetic. In some embodiments, the human is prediabetic. In some embodiments, the cell is from the human.

[0014] The present invention also provides methods of diagnosing an individual who has Type 2 diabetes or is prediabetic. In some embodiments, the method comprises, detecting in a sample from the individual the level of a polypeptide or the level of a polynucleotide encoding the polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14, wherein a modulated level of the polypeptide or polynucleotide in the sample compared to a level of the polypeptide or polynucleotide in either a lean individual or a previous sample from the individual indicates that the individual is diabetic or prediabetic.

[0015] In some embodiments, the detecting step comprises contacting the sample with an antibody that specifically binds to the polypeptide.

[0016] In some embodiments, the detecting step comprises quantifying mRNA encoding the polypeptide. In some embodiments, the mRNA is reverse transcribed and amplified in a polymerase chain reaction.

[0017] In some embodiments, the sample is a blood, urine or tissue sample.

[0018] The present invention also provides isolated nucleic acids that hybridize under stringent conditions to a polynucleotide encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14.

[0019] In some embodiments, the nucleic acid is selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13

[0020] The present invention also provides an expression cassette comprising a heterologous promoter operably linked to a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14.

[0021] The present invention also provides host cells transfected with a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14. In some embodiments, the host cell is a human cell. In some embodiments, the host cell is a bacterium.

[0022] The present invention also provides isolated polypeptides comprising an amino acid sequence at least 70% identical to SEQ ID NO: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14. In some embodiments, the polypeptide is SEQ ID NO: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14.

DEFINITIONS

[0023] “Insulin sensitivity” refers to the ability of a cell or tissue to respond to insulin. Responses include, e.g., glucose uptake of a cell or tissue in response to insulin stimulation. Sensitivity can be determined at an organismal, tissue or cellular level. For example, blood or urine glucose levels following a glucose tolerance test are indicative of insulin sensitivity. Other methods of measuring insulin sensitivity include, e.g., measuring glucose uptake (see, e.g., Garcia de Herreros, A., and Birnbaum, **[text missing or illegible when filed]**, *J. J. Biol. Chem.* 264, 19994-19999 (1989); Klip, A., Li, G., and Logan, W. *J. Am. J. Physiol.* 247, E291-296 (1984)), measuring the glucose infusion rate (GINF) into tissue such as the skeletal muscle (see, e.g., Ludvik et al., *J. Clin. Invest.* 100:2354 (1997); Frias et al., *Diabetes Care* 23:64, (2000)) and measuring sensitivity of GLUT4 translocation (e.g., as described herein) in response to insulin.

[0024] “Activity” of a polypeptide of the invention refers to structural, regulatory, or biochemical functions of a polypeptide in its native cell or tissue. Examples of activity of a polypeptide include both direct activities and indirect activities. Exemplary direct activities are the result of direct interaction with the polypeptide, e.g., enzymatic activity, ligand binding, production or depletion of second messengers (e.g., cAMP, cGMP, IP₃, DAG, or Ca²⁺), ion flux, phosphorylation levels, transcription levels, and the like. Exemplary indirect activities are observed as a change in phenotype or response in a cell or tissue to a polypeptide’s directed activity, e.g., modulating insulin sensitivity of a cell as a result of the interaction of the polypeptide with other cellular or tissue components.

[0025] “Predisposition for diabetes” occurs in a person when the person is at high risk for developing diabetes. A number of risk factors are known to those of skill in the art and include: genetic factors (e.g., carrying alleles that result in a higher occurrence of diabetes than in the average population or having parents or siblings with diabetes); overweight (e.g., body mass index (BMI) greater or equal to 25 kg/m³); habitual physical inactivity, race/ethnicity (e.g., African-American, Hispanic-American, Native Americans, Asian-Americans, Pacific Islanders); previously identified impaired fasting glucose or impaired glucose tolerance, hypertension (e.g., greater or equal to 140/90 mmHg in adults); HDL cholesterol less than or equal to 35 mg/dl; triglyceride levels greater or equal to 250 mg/dl; a history of gestational diabetes or delivery of a baby over nine pounds; and/or polycystic ovary syndrome. See, e.g., “Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus” and “Screening for Diabetes” *Diabetes Care* 25(1): S5-S24 (2002).

[0026] A “lean individual,” when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level less than 110 mg/dl or a 2 hour PG reading of 140 mg/dl. “Fasting” refers to no caloric intake for at least 8 hours. A “2 hour PG” refers to the level of blood glucose after challenging a patient to a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water. The overall test is generally referred to as an oral glucose tolerance test (OGTT). See, e.g., *Diabetes Care, Supplement* 2002, American Diabetes Association: Clinical Practice Recommendations 2002. The level of a polypeptide in a lean individual can be a reading from a single individual, but is

typically a statistically relevant average from a group of lean individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[0027] A “pre-diabetic individual,” when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level greater than 110 mg/dl but less than 126 mg/dl or a 2 hour PG reading of greater than 140 mg/dl but less than 200mg/dl. A “diabetic individual,” when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level greater than 126 mg/dl or a 2 hour PG reading of greater than 200 mg/dl.

[0028] A “diabetes-related nucleic acid” or “diabetes-related polynucleotide” (also referred to as a “nucleic acid of the invention” or a “polynucleotide of the invention”) of the invention is a subsequence or full-length polynucleotide sequence of a gene that encodes a polypeptide, whose activity modulates diabetes or insulin sensitivity, or whose presence or absence is indicative of diabetes or altered insulin sensitivity. Exemplary nucleic acids of the invention include those sequences substantially identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13 or encode polypeptides substantially identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO: 14.

[0029] An “agonist” refers to an agent that binds to, stimulates, increases, activates, facilitates, enhances activation, sensitizes or up regulates the activity or expression of a polypeptide of the invention.

[0030] An “antagonist” refers to an agent that binds to, partially or totally blocks stimulation, decreases, prevents, delays activation, inactivates, desensitizes, or down regulates the activity or expression of a polypeptide of the invention.

[0031] “Antibody” refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

[0032] An exemplary globulin (antibody, structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one “light” (about 25 kD) and one “heavy” chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

[0033] Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The

$F(ab)_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)_2$ dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (see, Paul (Ed.) *Fundamental Immunology*, Third Edition, Raven Press, NY (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant DNA methodologies (e.g., single chain Fv).

[0034] The terms “peptidomimetic” and “mimetic” refer to a synthetic chemical compound that has substantially the same structural and functional characteristics of the antagonists or agonists of the invention. Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compound are termed “peptide mimetics” or “peptidomimetics” (Fauchere, *J. Adv. Drug Res.* 15:29 (1986); Veber and Freidinger *TINS* p. 392 (1985); and Evans et al. *J. Med. Chem.* 30:1229 (1987), which are incorporated herein by reference). Peptide mimetics that are structurally similar to therapeutically useful peptides may be used to produce an equivalent or enhanced therapeutic or prophylactic effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (i.e., a polypeptide that has a biological or pharmacological activity), such as apolypeptide exemplified in this application, but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of, e.g., $-CH_2NH-$, $-CH_2S-$, $-CH_2-CH_2-$, $-CH=CH-$ (cis and trans), $-COCH_2-$, $-CH(OH)CH_2-$, and $-CH_2SO-$. The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetic's structure and/or activity. For example, a mimetic composition is within the scope of the invention if it is capable of carrying out the binding or other activities of an agonist or antagonist of a polypeptide of the invention.

[0035] The term “gene” means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

[0036] The term “isolated,” when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames that flank the gene and encode a protein other than the gene

of interest. The term “purified” denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

[0037] The term “nucleic acid” or “polynucleotide” refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka et al., *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol et al. (1992); Rossolini et al., *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

[0038] The terms “polypeptide,” “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins (i.e., antigens), wherein the amino acid residues are linked by covalent peptide bonds.

[0039] The term “amino acid” refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. “Amino acid mimetics” refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but which functions in a manner similar to a naturally occurring amino acid.

[0040] Amino acids may be referred to herein by either the commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0041] “Conservatively modified variants” applies to both amino acid and nucleic acid sequences. With respect to

particular nucleic acid sequences, “conservatively modified variants” refers to those nucleic acids that encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are “silent variations,” which are one species of conservatively modified variations. Every nucleic acid sequence herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid that encodes a polypeptide is implicit in each described sequence.

[0042] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

[0043] The following eight groups each contain amino acids that are conservative substitutions for one another:

[0044] 1) Alanine (A), Glycine (G);

[0045] 2) Aspartic acid (D), Glutamic acid (E);

[0046] 3) Asparagine (N), Glutamine (Q);

[0047] 4) Arginine (R), Lysine (K);

[0048] 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);

[0049] 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);

[0050] 7) Serine (S), Threonine (T); and

[0051] 8) Cysteine (C), Methionine (M)

[0052] (see, e.g., Creighton, *Proteins* (1984)).

[0053] “Percentage of sequence identity” is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (e.g., a polypeptide of the invention), which does not comprise additions or deletions, for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, divid-

ing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[0054] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same sequences are substantially identical if two sequences have a specified percentage of amino acid residues or nucleotides that are the same (i.e., 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region, or, when not specified, over the entire sequence), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. The invention provides polypeptides or polynucleotides that are substantially identical to the polypeptides or polynucleotides, respectively, exemplified herein (e.g., SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13 or SEQ ID NO:14). This definition also refers to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 nucleotides in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides in length.

[0055] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0056] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology* (1995 supplement)).

[0057] Two examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are

described in Altschul et al. (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

[0058] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0059] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

[0060] The phrase “selectively (or specifically) hybridizes to” refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

[0061] The phrase “stringent hybridization conditions” refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Probes*, “Overview of principles of hybridization and the strategy of nucleic acid assays” (1993). Generally, stringent conditions are selected to be about 5-10° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5×SSC, and 1% SDS, incubating at 42° C., or 5×SSC, 1% SDS, incubating at 65° C., with wash in 0.2×SSC, and 0.1% SDS at 55° C., 60° C., or 65° C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

[0062] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary “moderately stringent hybridization conditions” include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37° C., and a wash in 1×SSC at 45° C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

[0063] The phrase “a nucleic acid sequence encoding” refers to a nucleic acid which contains sequence information for a structural RNA such as rRNA, a tRNA, or the primary amino acid sequence of a specific protein or peptide, or a binding site for a transacting regulatory agent. This phrase specifically encompasses degenerate codons (i.e., different codons which encode a single amino acid) of the native

sequence or sequences that may be introduced to conform with codon preference in a specific host cell.

[0064] The term “recombinant” when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

[0065] The term “heterologous” when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequence unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

[0066] An “expression vector” is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

[0067] The phrase “specifically (or selectively) binds to an antibody” or “specifically (or selectively) immunoreactive with”, when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised against a protein having an amino acid sequence encoded by any of the polynucleotides of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins, except for polymorphic variants. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See, Harlow and Lane *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, NY (1988) for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically, a specific or selective reaction will be at least twice the background signal or noise and more typically more than 10 to 100 times background.

[0068] “Inhibitors,” “activators,” and “modulators” of expression or of activity are used to refer to inhibitory, activating, or modulating molecules, respectively, identified using in vitro and in vivo assays for expression or activity.

Modulators encompass e.g., ligands, agonists, antagonists, and their homologs and mimetics. The term "modulator" includes inhibitors and activators. Inhibitors are agents that, e.g., inhibit expression of a polypeptide of the invention or bind to, partially or totally block stimulation, decrease, prevent, delay activation, inactivate, desensitize, or down regulate [text missing or illegible when filed] activity of a polypeptide of the invention, e.g., antagonists. Activators are agents that, e.g., induce or activate the expression of a polypeptide of the invention or bind to, stimulate, increase, open, activate, facilitate, or enhance activation, sensitize or up regulate the activity of a polypeptide of the invention, e.g., agonists. Modulators include naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Such assays for inhibitors and activators include, e.g., applying putative modulator compounds to cells expressing a polypeptide of the invention and then determining the functional effects on a polypeptide of the invention activity, as described above. Samples or assays comprising a polypeptide of the invention that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of effect. Control samples (untreated with modulators) are assigned a relative activity value of 100%. Inhibition of a polypeptide of the invention is achieved when the polypeptide activity value relative to the control is about 80%, optionally 50% or 25, 10%, 5% or 1%. Activation of the polypeptide is achieved when the polypeptide activity value relative to the control is 110%, optionally 150%, optionally 200, 300%, 400%, 500%, or 1000-3000% or more higher.

DETAILED DESCRIPTION OF THE INVENTION

[0069] I. Introduction

[0070] The present invention demonstrates that, surprisingly, modulated levels of mRNA comprising sequences of the invention occur in muscle from diabetic individuals treated with thiazolidinedione (TZD) compared to levels of the mRNA in the individuals before TZD treatment. Without intending to limit the invention to a particular mechanism of action, it is believed that modulation of the expression or activity of the polypeptides or polynucleotides of the invention is beneficial in treating diabetic, pre-diabetic or obese insulin resistant, non-diabetic patients. Furthermore, modulated levels of the polypeptides of the invention are indicative of insulin resistance. Thus, the detection of a polypeptide of the invention is useful for diagnosis of diabetes and insulin resistance.

[0071] This invention also provides methods of using polypeptides of the invention and modulators of the polypeptides of the invention to diagnose and treat diabetes, pre-diabetes (including insulin resistant individuals) and related metabolic diseases. The present method also provides methods of identifying modulators of expression or activity of the polypeptides of the invention. Such modulators are useful for treating Type 2 diabetes as well as the pathological aspects of diabetes (e.g., insulin resistance).

II. General Recombinant Nucleic Acid Methods for Use with the Invention

[0072] In numerous embodiments of the present invention, nucleic acids encoding a polypeptide of the present invention will be isolated and cloned using recombinant methods. Such embodiments are used, e.g., to isolate polynucleotides

identical or substantially identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13 for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from an polypeptide or polynucleotide of the invention, to monitor gene expression, for the isolation or detection of sequences in different species, for diagnostic purposes in a patient, e.g., to detect mutations in a polypeptide or polynucleotide of the invention or to detect expression levels of nucleic acids or polypeptides. In some embodiments, the sequences encoding the polypeptides of the invention are operably linked to a heterologous promoter. In one embodiment, the nucleic acids of the invention are from any mammal, including, in particular, e.g., a human, a mouse, a rat, etc.

A. General Recombinant Nucleic Acid Methods

[0073] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook et al., *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel et al., eds., 1994).

[0074] For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa) or amino acid residue numbers. Proteins sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

[0075] Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter et. al., *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson & Reanier, *J. Chrom.* 255:137-149 (1983).

[0076] The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace et al., *Gene* 16:21-26 (1981).

B. Cloning Methods for the Isolation of Nucleotide Sequences Encoding Desired Proteins

[0077] In general, the nucleic acids encoding the subject proteins are cloned from DNA sequence libraries that are made to encode cDNA or genomic DNA. The particular sequences can be located by hybridizing with an oligonucleotide probe, the sequence of which can be derived from the sequences disclosed herein, which provide a reference for PCR primers and defines suitable regions for isolating probes specific for the polypeptides or polynucleotides of the invention. Alternatively, where the sequence is cloned into an expression library, the expressed recombinant protein can be detected immunologically with antisera or purified antibodies made against a polypeptide of interest, including those disclosed herein.

[0078] Methods for making and screening genomic and cDNA libraries are well known to those of skill in the art (see, e.g., Gubler and Hoffman *Gene* 25:263-269 (1983); Benton and Davis *Science*, 196:180-182 (1977); and Sambrook, supra).

[0079] Briefly, to make the cDNA library, one should choose a source that is rich in mRNA. The mRNA can then be made into cDNA, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. For a genomic library, the DNA is extracted from a suitable tissue and either mechanically sheared or enzymatically digested to yield fragments of preferably about 5-100 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged in vitro, and the recombinant phages are analyzed by plaque hybridization. Colony hybridization is carried out as generally described in Grunstein et al., *Proc. Natl. Acad. Sci. USA.*, 72:3961-3965 (1975).

[0080] An alternative method combines the use of synthetic oligonucleotide primers with polymerase extension on an mRNA or DNA template. Suitable primers can be designed from specific sequences disclosed herein. This polymerase chain reaction (PCR) method amplifies the nucleic acids encoding the protein of interest directly from mRNA, cDNA, genomic libraries or cDNA libraries. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other in vitro amplification methods may also be useful, for example, to clone nucleic acids encoding specific proteins and express said protein, to synthesize nucleic acids that will be used as probes for detecting the presence of mRNA encoding a polypeptide of the invention in physiological samples, for nucleic acid sequencing, or for other purposes (see, U.S. Pat. Nos. 4,683,195 and 4,683,202). Genes amplified by a PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

[0081] Appropriate primers and probes for identifying the genes encoding a polypeptide of the invention from mammalian tissues can be derived from the sequences provided herein. For a general overview of PCR, see, Innis et al. *PCR Protocols: A Guide to Methods and Applications*, Academic Press, San Diego (1990).

[0082] Synthetic oligonucleotides can be used to construct genes. This is done using a series of overlapping oligonucleotides, usually 40-120 bp in length, representing both the sense and anti-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned.

[0083] A polynucleotide encoding a polypeptide of the invention can be cloned using intermediate vectors before transformation into mammalian cells for expression. These intermediate vectors are typically prokaryote vectors or shuttle vectors. The proteins can be expressed in either prokaryotes or eukaryotes, using standard methods well known to those of skill in the art.

III. Purification of Proteins of the Invention

[0084] Either naturally occurring or recombinant polypeptides of the invention can be purified for use in functional assays. Naturally occurring polypeptides of the invention can be purified from any source (e.g., tissues of an organism

expressing an ortholog). Recombinant polypeptides can be purified from any suitable expression system.

[0085] The polypeptides of the invention may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; column chromatography, immunopurification methods, and others (see, e.g., Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Pat. No. 4,673,641; Ausubel et al., supra; and Sambrook et al., supra).

[0086] A number of procedures can be employed when recombinant polypeptides are being purified. For example, proteins having established molecular adhesion properties can be reversibly fused to a polypeptide of the invention. With the appropriate ligand, either protein can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein may be then removed by enzymatic activity. Finally polypeptides can be purified using immunoaffinity columns.

A. Purification of Proteins from Recombinant Bacteria

[0087] When recombinant proteins are expressed by the transformed bacteria in large amounts, typically after promoter induction, although expression can be constitutive, the proteins may form insoluble aggregates. There are several protocols that are suitable for purification of protein inclusion bodies. For example, purification of aggregate proteins (hereinafter referred to as inclusion bodies) typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells typically, but not limited to, by incubation in a buffer of about 100-150 µg/ml lysozyme and 0.1% Nonidet P40, a non-ionic detergent. The cell suspension can be ground using a Polytron grinder (Bran Instruments, Westbury, N.Y.). Alternatively, the cells can be sonicated on ice. Alternate methods of lysing bacteria are described in Ausubel et al. and Sambrook et al., both supra, and will be apparent to those of skill in the art.

[0088] The cell suspension is generally centrifuged and the pellet containing the inclusion bodies resuspended in buffer which does not dissolve but washes the inclusion bodies, e.g., 20 mM Tris-HCl (pH 7.2), 1 mM EDTA, 150 mM NaCl and 2% Triton-X 100, a non-ionic detergent. It may be necessary to repeat the wash step to remove as much cellular debris as possible. The remaining pellet of inclusion bodies may be resuspended in an appropriate buffer (e.g., 20 mM sodium phosphate, pH 6.8, 150 mM NaCl). Other appropriate buffers will be apparent to those of skill in the art.

[0089] Following the washing step, the inclusion bodies are solubilized by the addition of a solvent that is both a strong hydrogen acceptor and a strong hydrogen donor (or a combination of solvents each having one of these properties). The proteins that formed the inclusion bodies may then be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to, urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents that are capable of solubilizing aggregate-forming proteins, such as SDS (sodium dodecyl sulfate) and 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine

hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant allowing re-formation of the immunologically and/or biologically active protein of interest. After solubilization, the protein can be separated from other bacterial proteins by standard separation techniques.

[0090] Alternatively, it is possible to purify proteins from bacteria periplasm. Where the protein is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to those of skill in the art (see, Ausubel et al, supra). To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM MgSO₄ and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

B. Purification of Proteins from Insect Cells

[0091] Proteins can also be purified from eukaryotic gene expression systems as described in, e.g., Fernandez and Hoeffler, *Gene Expression Systems* (1999). In some embodiments, baculovirus expression systems are used to isolate proteins of the invention. Recombinant baculoviruses are generally generated by replacing the polyhedrin coding sequence of a baculovirus with a gene to be expressed (e.g., encoding a polypeptide of the invention). Viruses lacking the polyhedrin gene have a unique plaque morphology making them easy to recognize. In some embodiments, a recombinant baculovirus is generated by first cloning a polynucleotide of interest into a transfer vector (e.g., a pUC based vector) such that the polynucleotide is operably linked to a polyhedrin promoter. The transfer vector is transfected with wildtype DNA into an insect cell (e.g., Sf9, Sf21 or BT1-TN-5B1-4 cells), resulting in homologous recombination and replacement of the polyhedrin gene in the wildtype viral DNA with the polynucleotide of interest. Virus can then be generated and plaque purified. Protein expression results upon viral infection of insect cells. Expressed proteins can be harvested from cell supernatant if secreted, or from cell lysates if intracellular. See, e.g., Ausubel et al. and Fernandez and Hoeffler, supra.

C. Standard Protein Separation Techniques For [text missing or illegible when filed]fying Proteins

[0092] 1. Solubility Fractionation

[0093] Often as an initial step, and if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol is to add saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This

will precipitate the most hydrophobic proteins. The precipitate is discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, through either dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

[0094] 2. Size Differential Filtration

[0095] Based on a calculated molecular weight, a protein of greater and lesser size can be isolated using ultrafiltration through membranes of different pore sizes (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

[0096] 3. Column Chromatography

[0097] The proteins of interest can also be separated from other proteins on the basis of their size, net surface charge, hydrophobicity and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art.

[0098] Immunoaffinity chromatography using antibodies raised to a variety of affinity tags such as hemagglutinin (HA), FLAG, Xpress, Myc, hexahistidine (His), glutathione S transferase (GST) and the like can be used to purify polypeptides. The His tag will also act as a chelating agent for certain metals (e.g., Ni) and [text missing or illegible when filed]s the metals can also be used to purify His-containing polypeptides. After purification, the tag is optionally removed by specific proteolytic cleavage.

[0099] It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., Pharmacia Biotech).

IV. Detection of Polynucleotides of the Invention

[0100] Those of skill in the art will recognize that detection of expression of polynucleotides and polypeptides of the invention has many uses. For example, as discussed herein, detection of levels of polynucleotides and polypeptides of the invention in a patient is useful for diagnosing diabetes or a predisposition for at least some of the pathological effects of diabetes. Moreover, detection of gene expression is useful to identify modulators of expression of polynucleotides and polypeptides of the invention.

[0101] A variety of methods of specific DNA and RNA measurement that use nucleic acid hybridization techniques are known to those of skill in the art (see, Sambrook, supra). Some methods involve an electrophoretic separation (e.g., Southern blot for detecting DNA, and Northern blot for detecting RNA), but measurement of DNA and RNA can also be carried out in the absence of electrophoretic separation.

ration (e.g., by dot blot). Southern blot of genomic DNA (e.g., from a human) can be used for screening for restriction fragment length polymorphism (RFLP) to detect the presence of a genetic disorder affecting a polypeptide of the invention.

[0102] The selection of a nucleic acid hybridization format is not critical. A variety of nucleic acid hybridization formats are known to those skilled in the art. For example, common formats include sandwich assays and competition or displacement assays. Hybridization techniques are generally described in Hames and Higgins *Nucleic Acid Hybridization, A Practical Approach*, IRL Press (1985); Gall and Pardue, *Proc. Natl. Acad. Sci. USA.*, 63:378-383 (1969); and John et al. *Nature*, 223:582-587 (1969).

[0103] Detection of a hybridization complex may require the binding of a signal-generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal. The binding of the signal generation complex is also readily amenable to accelerations by exposure to ultrasonic energy.

[0104] The label may also allow indirect detection of the hybridization complex. For example, where the label is a hapten or antigen, the sample can be detected by using antibodies. In these systems, a signal is generated by attaching fluorescent or enzyme molecules to the antibodies or in some cases, by attachment to a radioactive label (see, e.g., Tijssen, "Practice and Theory of Enzyme Immunoassays," *Laboratory Techniques in Biochemistry and Molecular Biology*, Burdon and van Knippenberg Eds., Elsevier (1985), pp. 9-20).

[0105] The probes are typically labeled either directly, as with isotopes, chromophores, lumiphores, chromogens, or indirectly, such as with biotin, to which a streptavidin complex may later bind. Thus, the detectable labels used in the assays of the present invention can be primary labels (where the label comprises an element that is detected directly or that produces a directly detectable element) or secondary labels (where the detected label binds to a primary label, e.g., as is common in immunological labeling). Typically, labeled signal nucleic acids are used to detect hybridization. Complementary nucleic acids or signal nucleic acids may be labeled by any one of several methods typically used to detect the presence of hybridized polynucleotides. The most common method of detection is the use of autoradiography with ^3H , ^{125}I , ^{35}S , ^{14}C , or ^{32}P -labeled probes or the like.

[0106] Other labels include, e.g., ligands that bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies that can serve as specific binding pair members for a labeled ligand. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden *Introduction to Immunocytochemistry*, 2nd ed., Springer Verlag, NY (1997); and in Haugland *Handbook of Fluorescent Probes and Research Chemicals*, a combined handbook and catalogue Published by Molecular Probes, Inc. (1996).

[0107] In general, a detector that monitors a particular probe or probe combination is used to detect the detection reagent label. Typical detectors include spectrophotometers,

phototubes and photodiodes, microscopes, scintillation counters, cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill in the art. Commonly, an optical image of a substrate comprising bound labeling moieties is digitized for subsequent computer analysis.

[0108] The amount of, for example, an RNA is measured by quantitating the amount of label fixed to the solid support by binding of the detection reagent. Typically, the presence of a modulator during incubation will increase or decrease the amount of label fixed to the solid support relative to a control incubation that does not comprise the modulator, or as compared to a baseline established for a particular reaction type. Means of detecting and quantitating labels are well known to those of skill in the art.

[0109] In some embodiments, the target nucleic acid or the probe is immobilized on a solid support. Solid supports suitable for use in the assays of the invention are known to those of skill in the art. As used herein, a solid support is a matrix of material in a substantially fixed arrangement.

[0110] A variety of automated solid-phase assay techniques are also appropriate. For instance, very large scale immobilized polymer arrays (VLSIPSTM), i.e. Gene Chips or microarrays, available from Affymetrix, Inc. in Santa Clara, Calif. can be used to detect changes in expression levels of a plurality of genes involved in the same regulatory pathways simultaneously. See, Tijssen, supra., Fodor et al. (1991) *Science*, 251: 767-777; Sheldon et al. (1993) *Clinical Chemistry* 39(4): 718-719, and Kozal et al. (1996) *Nature Medicine* 2(7): 753-759. Similarly, spotted cDNA arrays (arrays of cDNA sequences bound to nylon, glass or another solid support) can also be used to monitor expression of a plurality of genes.

[0111] Typically, the array elements are organized in an ordered fashion so that each element is present at a specified location on the substrate. Because the array elements are at specified locations on the substrate, the hybridization patterns and intensities (which together create a unique expression profile) can be interpreted in terms of expression levels of particular genes and can be correlated with a particular disease or condition or treatment. See, e.g., Schena et al., *Science* 270: 467470 (1995)) and (Lockhart et al., *Nature Biotech.* 14: 1675-1680 (1996)).

[0112] Hybridization specificity can be evaluated by comparing the hybridization of specificity-control polynucleotide sequences to specificity-control polynucleotide probes that are added to a sample in a known amount. The specificity-control target polynucleotides may have one or more sequence mismatches compared with the corresponding polynucleotide sequences. In this manner, whether only complementary target polynucleotides are hybridizing to the polynucleotide sequences or whether mismatched hybrid duplexes are forming is determined.

[0113] Hybridization reactions can be performed in absolute or differential hybridization formats. In the absolute hybridization format, polynucleotide probes from one sample are hybridized to the sequences in a microarray format and signals detected after hybridization complex formation correlate to polynucleotide probe levels in a sample. In the differential hybridization format, the differ-

ential expression of a set of genes in two biological samples is analyzed. For differential hybridization, polynucleotide probes from both biological samples are prepared and labeled with different labeling moieties. A mixture of the two labeled polynucleotide probes is added to a microarray. The microarray is then examined under conditions in which the emissions from the two different labels are individually detectable. Sequences in the microarray that are hybridized to substantially equal numbers of polynucleotide probes derived from both biological samples give a distinct combined fluorescence (Shalon et al. PCT publication WO95/35505). In some embodiments, the labels are fluorescent labels with distinguishable emission spectra, such as Cy3 and Cy5 fluorophores.

[0114] After hybridization, the microarray is washed to remove nonhybridized nucleic acids and complex formation between the hybridizable array elements and the polynucleotide probes is detected. Methods for detecting complex formation are well known to those skilled in the art. In some embodiments, the polynucleotide probes are labeled with a fluorescent label and measurement of levels and patterns of fluorescence indicative of complex formation is accomplished by fluorescence microscopy, such as confocal fluorescence microscopy.

[0115] In a differential hybridization experiment, polynucleotide probes from two or more different biological samples are labeled with two or more different fluorescent labels with different emission wavelengths. Fluorescent signals are detected separately with different photomultipliers set to detect specific wavelengths. The relative abundances/expression levels of the polynucleotide probes in two or more samples are obtained.

[0116] Typically, microarray fluorescence intensities can be normalized to take into account variations in hybridization intensities when more than one microarray is used under similar test conditions. In some embodiments, individual polynucleotide probe/target complex hybridization intensities are normalized using the intensities derived from internal normalization controls contained on each microarray.

[0117] Detection of nucleic acids can also be accomplished, for example, by using a labeled detection moiety that binds specifically to duplex nucleic acids (e.g., an antibody that is specific for RNA-DNA duplexes). One example uses an antibody that recognizes DNA-RNA heteroduplexes in which the antibody is linked to an enzyme (typically by recombinant or covalent chemical bonding). The antibody is detected when the enzyme reacts with its substrate, producing a detectable product. [text missing or illegible when filed]utlee et al. (1989) *Analytical Biochemistry* 181:153-162; Bogulavski (1986) et al. *J. Immunol. Methods* 89:123-130; Prooijen-Knegt (1982) *Exp. Cell Res.* 141:397-407; Rudkin (1976) *Nature* 265:472-473, Stollar (1970) *PNAS* 65:993-1000; Ballard (1982) *Mol. Immunol.* 19:793-799; Pisetsky and Caster (1982) *Mol. Immunol.* 19:645-650; Viscidi et al. (1988) *J. Clin. Microbiol.* 41:199-209; and Kiney et al. (1989) *J. Clin. Microbiol.* 27:6-12 describe antibodies to RNA duplexes, including homo and heteroduplexes. Kits comprising antibodies specific for DNA:RNA hybrids are available, e.g., from Digene Diagnostics, Inc. (Beltsville, Md.).

[0118] In addition to available antibodies, one of skill in the art can easily make antibodies specific for nucleic acid duplexes using existing techniques, or modify those anti-

bodies that are commercially or publicly available. In addition to the art referenced above, general methods for producing polyclonal and monoclonal antibodies are known to those of skill in the art (see, e.g., Paul (ed) *Fundamental Immunology, Third Edition* Raven Press, Ltd., NY (1993); Coligan *Current Protocols in Immunology* Wiley/Greene, NY (1991); Harlow and Lane *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY (1989); Stites et al. (eds.) *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, Calif., and references cited therein; Goding *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, N.Y., (1986); and Kohler and Milstein *Nature* 256: 495-497 (1975)). Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors (see, Huse et al. *Science* 246:1275-1281 (1989); and Ward et al. *Nature* 341:544-546 (1989)). Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 μ M, preferably at least about 0.01 μ M or better, and most typically and preferably, 0.001 μ M or better.

[0119] The nucleic acids used in this invention can be either positive or negative probes. Positive probes bind to their targets and the presence of duplex formation is evidence of the presence of the target. Negative probes fail to bind to the suspect target and the absence of duplex formation is evidence of the presence of the target. For example, the use of a wild type specific nucleic acid probe or PCR primers may serve as a negative probe in an assay sample where only the nucleotide sequence of interest is present.

[0120] The sensitivity of the hybridization assays may be enhanced through use of a nucleic acid amplification system that multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBA, Canguene, Massasauga, Ontario) and Q Beta Replicase systems. These systems can be used to directly identify mutants where the PCR or LCR primers are designed to be extended or ligated only when a selected sequence is present. Alternatively, the selected sequences can be generally amplified using, for example, nonspecific PCR primers and the amplified target region later probed for a specific sequence indicative of a mutation. It is understood that various detection probes, including Taqman and molecular beacon probes can be used to monitor amplification reaction products, e.g., in real time.

[0121] An alternative means for determining the level of expression of the nucleic acids of the present invention is in situ hybridization. In situ hybridization assays are well known and are generally described in Angerer et al., *Methods Enzymol.* 152:649-660 (1987). In an in situ hybridization assay, cells, preferentially human cells from the cerebellum or the hippocampus, are fixed to a solid support, typically a glass slide. If DNA is to be probed, the cells are denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of specific probes that are labeled. The probes are preferably labeled with radioisotopes or fluorescent reporters.

[0122] Single nucleotide polymorphism (SNP) analysis is also useful for detecting differences between alleles of the

polynucleotides (e.g., genes) of the invention. SNPs linked to genes encoding polypeptides of the invention are useful, for instance, for diagnosis of diseases (e.g., diabetes) whose occurrence is linked to the gene sequences of the invention. For example, if an individual carries at least one SNP linked to a disease-associated allele of the gene sequences of the invention, the individual is likely predisposed for one or more of those diseases. If the individual is homozygous for a disease-linked SNP, the individual is particularly predisposed for occurrence of that disease (e.g., diabetes). In some embodiments, the SNP associated with the gene sequences of the invention is located within 300,000; 200,000; 100,000; 75,000; 50,000; or 10,000 base pairs from the gene sequence.

[0123] Various real-time PCR methods including, e.g., Taqman or molecular beacon-based assays (e.g., U.S. Pat. Nos. 5,210,015; 5,487,972; Tyagi et al., *Nature Biotechnology* 14:303 (1996); and PCT WO 95/13399 are useful to monitor for the presence of absence of a SNP. Additional SNP detection methods include, e.g., DNA sequencing, sequencing by hybridization, dot blotting, oligonucleotide array (DNA Chip) hybridization analysis, or are described in, e.g., U.S. Pat. No. 6,177,249; Landegren et al., *Genome Research*, 8:769-776 (1998); Botstein et al., *Am J Human Genetics* 32:314-331 (1980); Meyers et al., *Methods in Enzymology* 155:501-527 (1987); Ke[**text missing or illegible when filed**] et al., *Trends in Genetics* 7:5 (1991); Myers et al., *Science* 230:1242-1246 (1985); and Kwok et al., *Genomics* 23:138-144 (1994).

V. Immunological Detection of Polypeptides of the Invention

[0124] In addition to the detection of polynucleotides of the invention and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect polypeptides of the invention. Immunoassays can be used to qualitatively or quantitatively analyze polypeptides of the invention. A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

A. Antibodies to Target Proteins or Other Immunogens

[0125] Methods for producing polyclonal and monoclonal antibodies that react specifically with a protein of interest or other immunogen are known to those of skill in the art (see, e.g., Coligan, supra; and Harlow and Lane, supra; Stites et al., supra and references cited therein; Goding, supra; and Kohler and Milstein *Nature*, 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors (see, Huse et al., supra; and Ward et al., supra). For example, in order to produce antisera for use in an immunoassay, the protein of interest or an antigenic fragment thereof, is isolated as described herein. For example, a recombinant protein is produced in a transformed cell line. An inbred strain of mice or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

[0126] Polyclonal sera are collected and titered against the immunogen in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid

support. Polyclonal antisera with a titer of 10^4 or greater are selected and tested for their crossreactivity against proteins other than the polypeptides of the invention or even other homologous proteins from other organisms, using a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 mM, more usually at least about 1 μ M, preferably at least about 0.1 μ M or better, and most preferably, 0.01 μ M or better.

[0127] A number of proteins of the invention comprising immunogens may be used to produce antibodies specifically or selectively reactive with the proteins of interest. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring protein may also be used either in pure or impure form. Synthetic peptides made using the protein sequences described herein may also be used as an immunogen for the production of antibodies to the protein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells and purified as generally described supra. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the protein.

[0128] Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to polypeptides of the invention. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (see, Harlow and Lane, supra).

[0129] Monoclonal antibodies may be obtained using various techniques familiar to those of skill in the art. Typically, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see, Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include, e.g., transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences that encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al., supra.

[0130] Once target immunogen-specific antibodies are available, the immunogen can be measured by a variety of immunoassay methods with qualitative and quantitative results available to the clinician. For a review of immunological and immunoassay procedures in general see, Stites, supra. Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Maggio *Enzyme Immunoassay*, CRC Press, Boca Raton, Fla. (1980); Tijssen, supra; and Harlow and Lane, supra.

[0131] Immunoassays to measure target proteins in a human sample may use a polyclonal antiserum that was raised to full-length polypeptides of the invention or a fragment thereof. This antiserum is selected to have low cross-reactivity against other proteins and any such cross-reactivity is removed by immunoabsorption prior to use in the immunoassay.

B. Immunological Binding Assays

[0132] In some embodiments, a protein of interest is detected and/or quantified using any of a number of well-known immunological binding assays (see, e.g., U.S. Pat. Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also Asai *Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Academic Press, Inc. NY (1993); Stites, supra. Immunological binding assays (or immunoassays) typically utilize a "capture agent" to specifically bind to and often immobilize the analyte (e.g., full-length polypeptides of the present invention, or antigenic subsequences thereof). The capture agent is a moiety that specifically binds to the analyte. The antibody may be produced by any of a number of means well known to those of skill in the art and as described above.

[0133] Immunoassays also often utilize a labeling agent to bind specifically to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/protein complex.

[0134] In a preferred embodiment, the labeling agent is a second antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second antibody can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

[0135] Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G, can also be used as the label agents. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (see, generally, Kronval, et al. *J. Immunol.*, 111:1401-1406 (1973); and Akerstrom, et al. *J. Immunol.*, 135:2589-2542 (1985)).

[0136] Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. The incubation time will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10° C. to 40° C.

[0137] 1. Non-Competitive Assay Formats

[0138] Immunoassays for detecting proteins or analytes of interest from tissue samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured protein or analyte is directly

measured. In one preferred "sandwich" assay, for example, the capture agent (e.g., antibodies specific for the polypeptides of the invention) can be bound directly to a solid substrate where it is immobilized. These immobilized antibodies then capture the polypeptide present in the test sample. The polypeptide of the invention thus immobilized is then bound by a labeling agent, such as a second labelled antibody specific for the polypeptide. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

[0139] 2. Competitive Assay Formats

[0140] In competitive assays, the amount of protein or analyte present in the sample is measured indirectly by measuring the amount of an added (exogenous) protein or analyte displaced (or competed away) from a specific capture agent (e.g., antibodies specific for a polypeptide of the invention) by the protein or analyte present in the sample. The amount of immunogen bound to the antibody is inversely proportional to the concentration of immunogen present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of analyte may be detected by providing a labeled analyte molecule. It is understood that labels can include, e.g., radioactive labels as well as peptide or other tags that can be recognized by detection reagents such as antibodies.

[0141] Immunoassays in the competitive binding format can be used for cross-reactivity determinations. For example, the protein encoded by the sequences described herein can be immobilized on a solid support. Proteins are added to the assay and compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to that of the protein encoded by any of the sequences described herein. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the considered proteins, e.g., distantly related homologs.

[0142] The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps a protein of the present invention, to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than 10 times the amount of the protein partially encoded by a sequence herein that is required, then the second protein is said to specifically bind to an antibody generated to an immunogen consisting of the target protein.

[0143] 3. Other Assay Formats

[0144] In some embodiments, western blot (immunoblot) analysis is used to detect and quantify the presence of a

polypeptide of the invention in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support (such as, e.g., a nitrocellulose filter, a nylon filter, or a derivatized nylon filter) and incubating the sample with the antibodies that specifically bind the protein of interest. For example, antibodies are selected that specifically bind to the polypeptides of the invention on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (e.g., labeled sheep anti-mouse antibodies) that specifically bind to the antibodies against the protein of interest.

[0145] Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (e.g., antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (see, Monroe et al. (1986) *Amer. Clin. Prod. Rev.* 5:34-41).

[0146] 4. Labels

[0147] The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most labels useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g., Dynabeads™), fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, 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 calorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads.

[0148] The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on the sensitivity required, the ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

[0149] Non-radioactive labels are often attached by indirect means. The molecules can also be conjugated directly to signal generating compounds, e.g., by conjugation with an enzyme or fluorescent compound. A variety of enzymes and fluorescent compounds can be used with the methods of the present invention and are well-known to those of skill in the art (for a review of various labeling or signal producing systems which may be used, see, e.g., U.S. Pat. No. 4,391, 904).

[0150] Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence

may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple calorimetric labels may be detected directly by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

[0151] Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need to be labeled and the presence of the target antibody is detected by simple visual inspection.

VI. Identification of Modulators of Polypeptides of the Invention

[0152] Modulators of a polypeptide of the invention, i.e. agonists or antagonists of a polypeptide's activity, or polypeptide's or polynucleotide's expression, are useful for treating a number of human diseases, including diabetes. For example, administration of modulators can be used to treat diabetic patients or prediabetic individuals to prevent progression, and therefore symptoms, associated with diabetes (including insulin resistance).

[0153] A. Agents that Modulate Polypeptides of the Invention

[0154] The agents tested as modulators of polypeptides of the invention can be any small chemical compound, or a biological entity, such as a protein, sugar, nucleic acid or lipid. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays). Modulators also include agents designed to reduce the level of mRNA encoding a polypeptide of the invention (e.g. antisense molecules, ribozymes, DNazymes, small inhibitory RNAs and the like) or the level of translation from an mRNA (e.g., translation blockers such as an antisense molecules that are complementary to translation start or other sequences on an mRNA molecule). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, Mo.), Aldrich (St. Louis, Mo.), Sigma-Aldrich (St. Louis, Mo.), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

[0155] In some embodiments, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity.

The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

[0156] A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[0157] Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Pat. No. 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton et al., *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (e.g., PCT Publication No. WO 91/19735), encoded peptides (e.g., PCT Publication WO 93/20242), random bio-oligomers (e.g., PCT Publication No. WO 92/00091), benzodiazepines (e.g., U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs et al., *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara et al., *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann et al., *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen et al., *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho et al., *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell et al., *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (see Ausubel, Berger and Sambrook, all supra), peptide nucleic acid libraries (see, e.g., U.S. Pat. No. 5,539,083), antibody libraries (see, e.g., Vaughn et al., *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al., *Science*, 274:1520-1522 (1996) and U.S. Pat. No. 5,593,853), small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN, January 18, page 33 (1993); isoprenoids, U.S. Pat. No. 5,569,588; thiazolidinones and metathiazanones, U.S. Pat. No. 5,549,974; pyrrolidines, U.S. Pat. Nos. 5,525,735 and 5,519,134;olino compounds, U.S. Pat. No. 5,506,337; benzodiazepines, U.S. Pat. No. 5,288,514, and the like).

[0158] Devices for the preparation of combinatorial libraries are commercially available (see, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.). In addition, numerous combinatorial libraries are themselves commercially available (see, e.g., ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, Mo., 3D Pharmaceuticals, Exton, Pa., Martek Biosciences, Columbia, Md., etc.).

B. Methods of Screening for Modulators of the Polypeptides of the Invention

[0159] A number of different screening protocols can be utilized to identify agents that modulate the level of expression or activity of a polynucleotide of a polypeptide of the

invention in cells, particularly mammalian cells, and especially human cells. In general terms, the screening methods involve screening a plurality of agents to identify an agent that modulates the activity of a polypeptide of the invention by, e.g., binding to the polypeptide, preventing an inhibitor or activator from binding to the polypeptide, increasing association of an inhibitor or activator with the polypeptide, or activating or inhibiting expression of the polypeptide.

[0160] Any cell expressing a full-length polypeptide of the invention or a fragment thereof can be used to identify modulators. In some embodiments, the cells are eukaryotic cells lines (e.g., CHO or HEK293) transformed to express a heterologous polypeptide of the invention. In some embodiments, a cell expressing an endogenous polypeptide of the invention is used in screens. In other embodiments, modulators are screened for their ability to effect insulin responses.

[0161] 1. Polypeptide Binding Assays

[0162] Preliminary screens can be conducted by screening for agents capable of binding to polypeptides of the invention, as at least some of the agents so identified are likely modulators of a polypeptide of the invention. Binding assays are also useful, e.g., for identifying endogenous proteins that interact with polypeptides of the invention. For example, antibodies, receptors or other molecules that bind polypeptides of the invention can be identified in binding assays.

[0163] Binding assays usually involve contacting polypeptide of the invention with one or more test agents and allowing sufficient time for the protein and test agents to form a binding complex. Any binding complexes formed can be detected using any of a number of established analytical techniques. Protein binding assays include, but are not limited to, methods that measure co-precipitation or co-migration on non-denaturing SDS-polyacrylamide gels, and co-migration on Western blots (see, e.g., Bennet, J. P. and Yamamura, H. I. (1985) "Neurotransmitter, Hormone or Drug Receptor Binding Methods," in *Neurotransmitter Receptor Binding* (Yamura, H. I., et al., eds.), pp. 61-89. Other binding assays involve the use of mass spectrometry or NMR techniques to identify molecules bound to a polypeptide of the invention or displacement of labeled substrates. The polypeptides of the invention utilized in such assays can be naturally expressed, cloned or synthesized.

[0164] In addition, mammalian or yeast two-hybrid approaches (see, e.g., Bartel, P. L. et al. *Methods Enzymol*, 254:241 (1995)) can be used to identify polypeptides or other molecules that interact or bind when expressed together in a host cell.

[0165] 2. Polypeptide Activity

[0166] The activity of polypeptides of the invention can be assessed using a variety of in vitro and in vivo assays to determine functional, chemical, and physical effects, e.g., measuring ligand binding (e.g., radioactive or otherwise labeled ligand binding), second messengers (e.g., cAMP, cGMP, IP₃, DAG, or Ca²⁺), ion flux, phosphorylation levels, transcription levels, and the like. Furthermore, such assays can be used to test for inhibitors and activators of the polypeptides of the invention. Modulators can also be genetically altered versions of polypeptides of the invention.

[0167] The polypeptide of the assay will be selected from a polypeptide with substantial identity to a sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12 or SEQ ID NO:14 or other conservatively modified variants thereof. Generally, the amino acid sequence identity will be at least 70%, optionally at least 85%, optionally at least 90-95% to the polypeptides exemplified herein. Optionally, the polypeptide of the assays will comprise a fragment of a polypeptide of the invention, such as an extracellular domain, transmembrane domain, cytoplasmic domain, ligand binding domain, subunit association domain, active site, and the like. Either a polypeptide of the invention or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.

[0168] Modulators of polypeptide activity are tested using either recombinant or naturally occurring polypeptides of the invention. The protein can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal, either recombinant or naturally occurring. For example, tissue slices, dissociated cells, e.g., from tissues expressing polypeptides of the invention, transformed cells, or membranes can be used. Modulation is tested using one of the in vitro or in vivo assays described herein.

[0169] Modulator binding to polypeptides of the invention, a domain, or chimeric protein can be tested in solution, in a bilayer membrane, attached to a solid phase, in a lipid monolayer, or in vesicles. Binding of a modulator can be tested using, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index), hydrodynamic (e.g., shape), chromatographic, or solubility properties.

[0170] Samples or assays that are treated with a potential modulator (e.g., a "test compound") are compared to control samples without the test compound, to examine the extent of modulation. Control samples (untreated with activators or inhibitors) are assigned a relative activity value of 100. Inhibition of the polypeptides of the invention is achieved when the activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of the polypeptides of the invention is achieved when the activity value relative to the control is 110%, optionally 150%, 200%, 300%, 400%, 500%, or 1000-2000%.

[0171] 3. Expression Assays

[0172] Screening for a compound that modulates the expression of a polynucleotide or a polypeptide of the invention is also provided. Screening methods generally involve conducting cell-based assays in which test compounds are contacted with one or more cells expressing a polynucleotide or a polypeptide of the invention, and then detecting an increase or decrease in expression (either transcript or translation product). Assays can be performed with any cells that express a polynucleotide or a polypeptide of the invention.

[0173] Expression can be detected in a number of different ways. As described infra, the expression level of a polynucleotide of the invention in a cell can be determined by probing the mRNA expressed in a cell with a probe that specifically hybridizes with a transcript (or complementary nucleic acid derived therefrom) of a polynucleotide of the

invention. Probing can be conducted by lysing the cells and conducting Northern blots or without lysing the cells using in situ-hybridization techniques. Alternatively, a polypeptide of the invention can be detected using immunological methods in which a cell lysate is probed with antibodies that specifically bind to the polypeptide.

[0174] The level of expression or activity of a polynucleotide or a polypeptide of the invention can be compared to a baseline value. The baseline value can be a value for a control sample or a statistical value that is representative of expression levels of a polynucleotide or a polypeptide of the invention for a control population (e.g., lean individuals as described herein) or cells (e.g., tissue culture cells not exposed to a modulator). Expression levels can also be determined for cells that do not express the polynucleotide or a polypeptide of the invention as a negative control. Such cells generally are otherwise substantially genetically the same as the test cells.

[0175] A variety of different types of cells can be utilized in the reporter assays. Cells that do not endogenously express a polypeptide of the invention can be prokaryotic, but are preferably eukaryotic. The eukaryotic cells can be any of the cells typically utilized in generating cells that harbor recombinant nucleic acid constructs. Exemplary eukaryotic cells include, but are not limited to, yeast, and various higher eukaryotic cells such as the HEK293, HepG2, COS, CHO and HeLa cell lines.

[0176] Various controls can be conducted to ensure that an observed activity is authentic including running parallel reactions with cells that lack the reporter construct or by not contacting a cell harboring the reporter construct with test compound. Compounds can also be further validated as described below.

[0177] 4. Validation

[0178] Agents that are initially identified by any of the foregoing screening methods can be further tested to validate the apparent activity. Modulators that are selected for further study can be tested on the "classic" insulin responsive cell line, mouse 3T3-L1 adipocytes, muscle cells such as L6 cells and the like. Cells (e.g., adipocytes or muscle cells) are pre-incubated with the modulators and tested for acute (up to 4 hours) and chronic (overnight) effects on basal and insulin-stimulated GLUT4 translocation and glucose uptake.

[0179] Following such studies, validity of the modulators is tested in suitable animal models. The basic format of such methods involves administering a lead compound identified during an initial screen to an animal that serves as a model for humans and then determining if expression of activity of a polypeptide of the invention is in fact modulated.

[0180] The effect of the compound will be assessed in either diabetic animals or in diet induced insulin resistant animals. The blood glucose and insulin levels will be determined. The animal models utilized in validation studies generally are mammals of any kind. Specific examples of suitable animals include, but are not limited to, primates, mice and rats. For example, monogenic models of diabetes (e.g., ob/ob and db/db mice, Zucker rats and Zucker Diabetic Fatty rats etc) or polygenic models of diabetes (e.g., OLETF rats, GK rats, NSY mice, and KK mice) can be useful for validating modulation of a polypeptide of the invention in a

diabetic or insulin resistant animal. In addition, transgenic animals expressing human polypeptides of the invention can be used to further validate drug candidates.

C. Solid Phase and Soluble High Throughput Assays

[0181] In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (e.g., 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100 to about 1500 different compounds. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 or more different compounds are possible using the integrated systems of the invention. In addition, microfluidic approaches to reagent manipulation can be used.

[0182] A molecule of interest (e.g., a polypeptide or polynucleotide of the invention, or a modulator thereof) can be bound to the solid-state component, directly or indirectly, via covalent or non-covalent linkage, e.g., via a tag. The tag can be any of a variety of components. In general, a molecule that binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest is attached to the solid support by interaction of the tag and the tag binder.

[0183] A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, poly-His, etc.) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders (see, SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis Mo.).

[0184] Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody that recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs, such as agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherin family, the integrin family, the selectin family, and the like; see, e.g., Pigott & Power, *The Adhesion Molecule Facts Book I* (1993)). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), intracellular receptors (e.g., which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

[0185] Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyl-

eneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

[0186] Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly-gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to those of skill in the art. For example, poly(ethylene glycol) linkers are available from Shearwater Polymers, Inc., Huntsville, Ala. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

[0187] Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent that fixes a chemical group to the surface that is reactive with a portion of the tag binder. For example, groups that are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature (see, e.g., Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, e.g., peptides); Geysen et al., *J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank and Doring, *Tetrahedron* 44:6031-6040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor et al., *Science*, 251:767-777 (1991); Sheldon et al., *Clinical Chemistry* 39(4):718-719 (1993); and Kozal et al., *Nature Medicine* 2(7):753-759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

[0188] The invention provides in vitro assays for identifying, in a high throughput format, compounds that can modulate the expression or activity of a polypeptide of the invention. Control reactions that measure activity of a polypeptide of the invention in a cell in a reaction that does not include a potential modulator are optional, as the assays are highly uniform. Such optional control reactions are appropriate and increase the reliability of the assay. Accordingly, in some embodiments, the methods of the invention include such a control reaction. For each of the assay formats described, "no modulator" control reactions that do not include a modulator provide a background level of binding activity.

[0189] In some assays it will be desirable to have positive controls. At least two types of positive controls are appropriate. First, a known activator of a polypeptide or a polynucleotide of the invention can be incubated with one sample of the assay, and the resulting increase in signal resulting from an increased expression level or activity of a polypeptide or a polynucleotide of the invention are determined according to the methods herein. Second, a known inhibitor of a polypeptide or a polynucleotide of the invention can be added, and the resulting decrease in signal for the expression or activity of a polypeptide or a polynucleotide of the invention can be similarly detected. It will be appreciated

that modulators can also be combined with activators or inhibitors to find modulators that inhibit the increase or decrease that is otherwise caused by the presence of the known modulator of a polypeptide or a polynucleotide of the invention.

VII. Compositions, Kits and Integrated Systems

[0190] The invention provides compositions, kits and integrated systems for practicing the assays described herein using nucleic acids or polypeptides of the invention, antibodies, etc.

[0191] The invention provides assay composition for use in solid phase assays; such compositions can include, for example, one or more nucleic acids encoding a polypeptide of the invention immobilized on a solid support, and a labeling reagent. In each case, the assay compositions can also include additional reagents that are desirable for hybridization. Modulators of expression or activity of a polypeptide of the invention can also be included in the assay compositions.

[0192] The invention also provides kits for carrying out the assays of the invention. The kits typically include a probe that comprises an antibody that specifically binds to a polypeptide of the invention or a polynucleotide sequence encoding such polypeptides, and a label for detecting the presence of the probe. The kits may include at least one polynucleotide sequence encoding a polypeptide of the invention. Kits can include any of the compositions noted above, and optionally further include additional components such as instructions to practice a high-throughput method of assaying for an effect on expression of the genes encoding a polypeptide of the invention, or on activity of a polypeptide of the invention, one or more containers or compartments (e.g., to hold the probe, labels, or the like), a control modulator of the expression or activity of a polypeptide of the invention, a robotic armature for mixing kit components or the like.

[0193] The invention also provides integrated systems for high-throughput screening of potential modulators for an effect on the expression or activity of a polypeptide of the invention. The systems can include a robotic armature which transfers fluid from a source to a destination, a controller which controls the robotic armature, a label detector, a data storage unit which records label detection, and an assay component such as a microtiter dish comprising a well having a reaction mixture or a substrate comprising a fixed nucleic acid or immobilization moiety.

[0194] A number of robotic fluid transfer systems are available, or can easily be made from existing components. For example, a Zymate XP (Zymark Corporation; Hopkinton, Mass.) automated robot using a Microlab 2200 (Hamilton; Reno, Nev.) pipetting station can be used to transfer parallel samples to 96 well microtiter plates to set up several parallel simultaneous binding assays.

[0195] Optical images viewed (and, optionally, recorded) by a camera or other recording device (e.g., a photodiode and data storage device) are optionally further processed in any of the embodiments herein, e.g., by digitizing the image and storing and analyzing the image on a computer. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video or digitized optical image.

[0196] One conventional system carries light from the specimen field to a cooled charge-coupled device (CCD) camera, in common use in the art. A CCD camera includes an array of picture elements (pixels). The light from the specimen is imaged on the CCD. Particular pixels corresponding to regions of the specimen (e.g., individual hybridization sites on an array of biological polymers) are sampled to obtain light intensity readings for each position. Multiple pixels are processed in parallel to increase speed. The apparatus and methods of the invention are easily used for viewing any sample, e.g., by fluorescent or dark field microscopic techniques.

VIII. Administration and Pharmaceutical Compositions

[0197] Modulators of the polypeptides of the invention (e.g., antagonists or agonists) can be administered directly to the mammalian subject for modulation of activity of a polypeptide of the invention in vivo. Administration is by any of the routes normally used for introducing a modulator compound into ultimate contact with the tissue to be treated and is well known to those of skill in the art. Although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0198] The pharmaceutical compositions of the invention may comprise a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there are a wide variety of suitable formulations of pharmaceutical compositions of the present invention (see, e.g., *Remington's Pharmaceutical Sciences*, 17th ed. 1985)).

[0199] The modulators (e.g., agonists or antagonists) of the expression or activity of the a polypeptide of the invention, alone or in combination with other suitable components, can be prepared for injection or for use in a pump device. Pump devices (also known as "insulin pumps") are commonly used to administer insulin to patients and therefore can be easily adapted to include compositions of the present invention. Manufacturers of insulin pumps include Animas, Disetronic and MiniMed.

[0200] The modulators (e.g., agonists or antagonists) of the expression or activity of a polypeptide of the invention, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[0201] Formulations suitable for administration include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, orally, nasally, topically, intravenously, intraperitoneally, or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Solutions and suspensions can be prepared from sterile

powders, granules, and tablets of the kind previously described. The modulators can also be administered as part of a prepared food or drug.

[0202] The dose administered to a patient, in the context of the present invention should be sufficient to induce a beneficial response in the subject over time. The optimal dose level for any patient will depend on a variety of factors including the efficacy of the specific modulator employed, the age, body weight, physical activity, and diet of the patient, on a possible combination with other drugs, and on the severity of the case of diabetes. It is recommended that the daily dosage of the modulator be determined for each individual patient by those skilled in the art in a similar way as for known insulin compositions. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

[0203] In determining the effective amount of the modulator to be administered a physician may evaluate circulating plasma levels of the modulator, modulator toxicity, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

[0204] For administration, modulators of the present invention can be administered at a rate determined by the LD-50 of the modulator, and the side-effects of the modulator at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

[0205] The compounds of the present invention can also be used effectively in combination with one or more additional active agents depending on the desired target therapy (see, e.g., Turner, N. et al. *Prog. Drug Res.* (1998) 51: 33-94; Haffner, S. *Diabetes Care* (1998) 21: 160-178; and DeFronzo, R. et al. (eds.), *Diabetes Reviews* (1997) Vol. 5 No. 4). A number of studies have investigated the benefits of combination therapies with oral agents (see, e.g., Mahler, R., *J. Clin. Endocrinol. Metab.* (1999) 84: 1165-71; United Kingdom Prospective Diabetes Study Group: UKPDS 28, *Diabetes Care* (1998) 21: 87-92; Bardin, C. W., (ed.), *Current Therapy In Endocrinology And Metabolism*, 6th Edition (Mosby-Year Book, Inc., St. Louis, Mo. 1997); Chiasson, J. et al., *Ann. Intern. Med.* (1994) 121: 928-935; Coniff, R. et al., *Clin. Ther.* (1997) 19: 16-26; Coniff, R. et al., *Am. J. Med.* (1995) 98: 443-451; and Iwamoto, Y. et al., *Diabet. Med.* (1996) 13 365-370; Kwiterovich, P. *Am. J. Cardiol* (1998) 82(12A): 3U-17U). These studies indicate that modulation of diabetes, among other diseases, can be further improved by the addition of a second agent to the therapeutic regimen. Combination therapy includes administration of a single pharmaceutical dosage formulation that contains a modulator of the invention and one or more additional active agents, as well as administration of a modulator and each active agent in its own separate pharmaceutical dosage formulation. For example, a modulator and a thiazolidinedione can be administered to the human subject together in a single oral dosage composition, such as a tablet or capsule, or each agent can be administered in separate oral dosage formulations. Where separate dosage formulations are used, a modulator and one or more additional active agents can be administered at essentially the same time (i.e., concurrently), or at separately staggered times (i.e., sequentially). Combination therapy is understood to include all these regimens.

[0206] One example of combination therapy can be seen in treating pre-diabetic individuals (e.g., to prevent progression into type 2 diabetes) or diabetic individuals (or treating diabetes and its related symptoms, complications, and disorders), wherein the modulators can be effectively used in combination with, for example, sulfonyleureas (such as chlorpropamide, tolbutamide, acetohexamide, tolazamide, glyburide, gliclazide, glynase, glimepiride, and glipizide); biguanides (such as metformin); a PPAR beta delta agonist; a ligand or agonist of PPAR gamma such as thiazolidinediones (such as ciglitazone, pioglitazone (see, e.g., U.S. Pat. No. 6,218,409), troglitazone, and rosiglitazone (see, e.g., U.S. Pat. No. 5,859,037)); PPAR alpha agonists such as clofibrate, gemfibrozil, fenofibrate, ciprofibrate, and bezafibrate; dehydroepiandrosterone (also referred to as DHEA or its conjugated sulphate ester, DHEA-SO₄); antigluco-corticoids; TNF α inhibitors; α -glucosidase inhibitors (such as acarbose, miglitol, and voglibose); amylin and amylin derivatives (such as pramlintide, (see, also, U.S. Pat. Nos. 5,902,726; 5,124,314; 5,175,145 and 6,143,718.)); insulin secretagogues (such as repaglinide, gliquidone, and nateglinide (see, also, U.S. Pat. Nos. 6,251,856; 6,251,865; 6,221,633; 6,174,856)), and insulin.

IX. Gene Therapy

[0207] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered polypeptides of the invention in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding polypeptides of the invention to cells in vitro. In some embodiments, the nucleic acids encoding polypeptides of the invention are administered for in vivo or ex vivo gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada et al., in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds) (1995); and Yu et al., *Gene Therapy* 1:13-26 (1994).

[0208] Methods of non-viral delivery of nucleic acids encoding engineered polypeptides of the invention include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. No. 5,049,386, U.S. Pat. No. 4,946,787; and U.S. Pat. No. 4,897,355) and lipofection reagents are sold commercially (e.g., TransfectamTM and LipofectinTM). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (ex vivo administration) or target tissues (in vivo administration).

[0209] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid com-

plexes, is well known to one of skill in the art (see, e.g., Crystal, *Science* 270:404-410 (1995); Blaese et al., *Cancer Gene Ther.* 2:291-297 (1995); Behr et al., *Bioconjugate Chem.* 5:382-389 (1994); Remy et al., *Bioconjugate Chem.* 5:647-654 (1994); Gao et al., *Gene Therapy* 2:710-722 (1995); Ahmad et al., *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0210] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered polypeptides of the invention take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro and the modified cells are administered to patients (ex vivo). Conventional viral based systems for the delivery of polypeptides of the invention could include retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0211] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Sommerfelt et al., *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63:2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0212] In applications where transient expression of the polypeptides of the invention is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,638; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994)). Construction of recombinant AAV vectors is described in a

number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulsd et al., *J. Virol.* 63:03822-3828 (1989).

[0213] pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar et al., *Blood* 85:3048-305 (1995); Kohn et al., *Nat. Med.* 1:1017-102 (1995); Malech et al., *PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese et al., *Science* 270:475-480 (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem et al., *Immunol Immunother.* 44(1):10-20 (1997); Dranoff et al., *Hum. Gene Ther.* 1:111-2 (1997)).

[0214] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner et al., *Lancet* 351:9117 1702-3 (1998); Kearns et al., *Gene Ther.* 9:748-55 (1996)).

[0215] Replication-deficient recombinant adenoviral vectors (Ad) can be engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiply types of tissues in vivo, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman et al., *Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker et al., *Infection* 24:15-10 (1996); Sterman et al., *Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh et al., *Hum. Gene Ther.* 2:205-18 (1995); Alvarez et al., *Hum. Gene Ther.* 5:597-613 (1997); Topf et al., *Gene Ther.* 5:507-513 (1998); Sterman et al., *Hum. Gene Ther.* 7:1083-1089 (1998).

[0216] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ 2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication

of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[0217] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han et al., *PNAS* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., Fab or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

[0218] Gene therapy vectors can be delivered in vivo by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells ex vivo, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0219] Ex vivo cell transfection for diagnostics, rich, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In some embodiments, cells are isolated from the subject organism, transfected with a nucleic acid (gene or cDNA) encoding a polypeptides of the invention, and re-infused back into the subject organism (e.g., patient). Various cell types suitable for ex vivo transfection are well known to those of skill in the art (see, e.g., Freshney et al., *Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

[0220] In one embodiment, stem cells are used in ex vivo procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types in vitro, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells in vitro into clinically important immune cell types using cytokines such as GM-CSF, IFN- γ and TNF- α are known (see Inaba et al., *J. Exp. Med.* 176:1693-1702 (1992)).

[0221] Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panb cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (see Inaba et al., *J. Exp. Med.* 176:1693-1702 (1992)).

[0222] Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic nucleic acids can be also administered directly to the organism for transduction of cells in vivo. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0223] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention, as described below (see, e.g., *Remington's Pharmaceutical Sciences*, 17th ed., 1989).

X. Diagnosis of Diabetes

[0224] The present invention also provides methods of diagnosing diabetes or a predisposition of at least some of the pathologies of diabetes. Diagnosis can involve determination of a genotype of an individual (e.g., with SNPs) and comparison of the genotype with alleles known to have an association with the occurrence of diabetes. Alternatively, diagnosis also involves determining the level of a polypeptide or polynucleotide of the invention in a patient and then comparing the level to a baseline or range. Typically, the baseline value is representative of a polypeptide or polynucleotide of the invention in a healthy (e.g., lean) person.

[0225] As discussed above, variation of levels (e.g., low or high levels) of a polypeptide or polynucleotide of the invention compared to the baseline range indicates that the patient is either diabetic or at risk of developing at least some of the pathologies of diabetes (e.g., pre-diabetic). The level of a polypeptide in a lean individual can be a reading from a single individual, but is typically a statistically relevant average from a group of lean individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[0226] In some embodiments, the level of polypeptide or polynucleotide of the invention is measured by taking a blood, urine or tissue sample from a patient and measuring the amount of a polypeptide or polynucleotide of the invention in the sample using any number of detection methods, such as those discussed herein. For instance, fasting and fed blood or urine levels can be tested.

[0227] In some embodiments, the baseline level and the level in a lean sample from an individual, or at least two samples from the same individual differ by at least about 5%, 10%, 20%, 50%, 75%, 100%, 150%, 200%, 300%, 400%, 500%, 1000% or more. In some embodiments, the

sample from the individual is greater by at least one of the above-listed percentages relative to the baseline level. In some embodiments, the sample from the individual is lower by at least one of the above-listed percentages relative to the baseline level.

[0228] In some embodiments, the level of a polypeptide or polynucleotide of the invention is used to monitor the effectiveness of antidiabetic therapies such as thiazolidinediones, metformin, sulfonylureas and other standard therapies. In some embodiments the activity or expression of a polypeptide or polynucleotide of the invention will be measured prior to and after treatment of diabetic or pre-diabetic patients with antidiabetic therapies as a surrogate marker of clinical effectiveness. For example, the greater the reduction in expression or activity of a polypeptide of the invention indicates greater effectiveness.

[0229] Glucose/insulin tolerance tests can also be used to detect the effect of glucose levels on levels of a polypeptide or polynucleotide of the invention. In glucose tolerance tests, the patient's ability to tolerate a standard oral glucose load is evaluated by assessing serum and urine specimens for glucose levels. Blood samples are taken before the glucose is ingested, glucose is given by mouth, and blood or urine glucose levels are tested at set intervals after glucose ingestion. Similarly, meal tolerance tests can also be used to detect the effect of insulin or food, respectively, on levels of a polypeptide or polynucleotide of the invention.

[0230] All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

[0231] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

[0232] The following examples are offered to illustrate, but not to limit the claimed invention.

[0233] The molecular mechanism by which thiazolidinediones (TZDs) cause an increase in peripheral insulin

sensitivity was studied. Genes in muscle or fat whose expression is altered by TZDs may lie on a pathway leading from TZD treatment to increased insulin sensitivity. Modulators of such genes can elicit the same effect as TZD treatment. Moreover, such modulators can lack some of the side effects of TZD. The majority of glucose disposal occurs in muscle. For this reason, gene expression profiling in human muscle from diabetics treated with troglitazone was used to identify genes important for TZD action and therefore treatment of diabetes and insulin resistance.

[0234] Gene expression profiling was performed on tissue samples (muscle) obtained from lean obese and diabetic individuals. Two studies were performed. In the first study, basal samples were isolated from all individuals at the beginning of a 5 hour hyperinsulinemic euglycemic clamp. Clamp samples were isolated the end of this procedure. Similar basal and clamp samples were taken 3 months later after all patients had taken the insulin sensitizing drug troglitazone.

[0235] In the second study, samples were obtained from lean, obese and diabetic individuals before and after a hyperinsulinemic euglycemic clamp. No troglitazone treatment was used. For all tissue samples mRNA was isolated from these muscle samples and converted to cRNA by standard procedures. The gene expression profile for each individual was determined by hybridization of cRNA to custom synthesized Affymetrix chips.

[0236] Gene expression profile differences were calculated as follows. The expression level of a particular gene is indicated by its 'average difference score'. The raw data was analyzed by a statistical test to remove 'outliers'. The mean 'average difference score' was then calculated from the average difference scores for all individuals in a particular treatment group. Genes were determined to be changed in condition 1 (such as basal diabetics before troglitazone (tro) treatment) versus condition 2 (such as basal diabetics after troglitazone treatment) by calculating the Students t test statistic between the two conditions and selecting those with t less than or equal to 0.05. Fold change was determined as the ratio of mean average difference score in condition 2 to the mean average difference score in condition 1.

[0237] Probe set mbxhummus 24330 detects expression of connective tissue growth factor (CTGF) mRNA.

[0238] Expression of transcripts encoding Connective Tissue Growth Factor (SEQ ID NO:2) was decreased in treated patients compared to untreated patients in this study.

B/C	Diabetic Pre-Trog			Diabetic Post-Trog			Fold Change	t test	Gene name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	409	83	9	29	5	10	0.07	0.016	Connective Tissue Growth Factor

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" and "Post-Trog" indicates sample was taken before or after 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetic post tro/diabetic pre tro.

- [0239] Connective Tissue Growth Factor (SEQ ID NO:2) contains the following protein domains:
- [0240] Signal peptide at amino acids 1 to 27;
- [0241] Insulin growth factor-binding protein homologues domain at amino acids 27 to 97;
- [0242] Von Willebrand factor (vWF) type C domain at amino acids 103 to 166;
- [0243] Thrombospondin type 1 repeat Domain at amino acids 200 to 243; and
- [0244] C-terminal cystine knot-like domain at amino acids 261 to 330.
- [0245] Probe set mbxhummus23808 detects expression of TIEG mRNA
- [0246] Expression of transcripts encoding TIEG (SEQ ID NO:8) was higher in diabetic patients compared to lean, non-diabetic patients in this study.

- [0251] zf-C2H2 domain encoded by amino acids 418 to 440.
- [0252] A splice variant of TIEG is also known (SEQ ID NO:13), which encodes the amino acid sequence displayed in SEQ ID NO:14.
- [0253] Transcripts encoding TIEG (SEQ ID NO:13) contain the following protein domains:
- [0254] zf-C2H2 domain encoded by amino acids 369 to 393
- [0255] zf-C2H2 domain encoded by amino acids 399 to 423
- [0256] zf-C2H2 domain encoded by amino acids 429 to 440.
- [0257] The gene expression changes observed in the human muscle samples derived from lean and diabetic individuals or from diabetic individuals post-Troglitazone

B/C	Lean Pre-Trog			Diabetic Pre-Trog			Fold Change	Students	
	Mean Expr	SEM	n	Mean Expr	SEM	n		t test	Gene name
B	1593	161	17	3335	249	19	2.094	0.000	TIEG

[0247] Expression of transcripts encoding TIEG (SEQ ID NO:8) was reduced in tro-treated patients compared to untreated patients.

treatment, were further confirmed by the use of real-time quantitative PCR. A combination of specific PCR primers and a Taqman probe were designed to detect and quantify

B/C	Diabetic Pre-Trog			Diabetic Post-Trog			Fold Change	Students	
	Mean Expr	SEM	n	Mean Expr	SEM	n		t test	Gene name
B	4061	559	9	2634	285	10	0.65	0.022	TIEG

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" and "Post-Trog" indicates sample was taken before or after 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetic post-trog/diabetic pre-trog.

Legend: B/C indicates sample is from Basal or Clamp; "Pre-Trog" and "Post-Trog" indicates sample was taken before or after 3 months of Troglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetic post-trog/diabetic pre-trog.

expression levels for each gene. Relative gene expression levels were determined in this manner in each patient and the mean expression level in either the lean or diabetic patient population was then calculated.

- [0248] Transcripts encoding TIEG (SEQ ID NO:8) contain the following protein domains:
- [0249] zf-C2H2 domain encoded by amino acids 358 to 382
- [0250] zf-C2H2 domain encoded by amino acids 388 to 412

[0258] In order to demonstrate the ability of the gene to modulate insulin sensitivity a number of different analyses were undertaken. One analysis called "Glut 4 translocation" involves over-expressing the gene in 3T3-L1 adipocytes and then monitoring the ability of insulin to cause the movement (translocation) of the major insulin-regulated glucose transporter, Glut 4, to the cell surface. The cDNA encoding the human gene was sub-cloned into the mammalian expression

vector pcDNA3.1 which provides for a V5 epitope tag. This plasmid encoding the human gene was then introduced into differentiated murine 3T3-L1 adipocytes along with a cloned Glut 4 expression construct using an electroporation procedure essentially as described in Kanzaki et al. *J. Biol. Chem* 275 7167-7175 (2000). A plasmid expressing LacZ was used as a negative control. After stimulation with increasing concentrations of insulin, cells co-expressing both the human gene and cloned Glut 4 were scored for the presence of Glut 4 at the cell surface using fluorescence microscopy.

[0259] In some instances another analysis called "glucose transport" was used. This involves over-expressing the gene in either differentiated 3T3-L1 adipocytes or differentiated L6 myotubes and then monitoring the ability of insulin to stimulate glucose transport into these two cell types. A recombinant adenovirus encoding the human gene tagged with a FLAG epitope tag was prepared essentially as described in Zhou et al. *Proc. Nat. Acad. Sci. USA*. 95: 2509-2514. 3T3-L1 adipocytes or L6 myotubes were infected with recombinant adenovims expressing the human gene. A virus expressing an irrelevant protein was used as the control. 24 h post-infection the cells are stimulated with increasing concentrations of insulin and glucose transport into the cells determined by the counting the accumulation of a radiolabelled glucose analog (2-deoxyglucose) essentially as describe in Kotani et al. *Mol Cell. Biol.* 18 6971-6982 (1998), Fujishiro et al. *J. Biol. Chem.* 276 19800-19806 (2001), Ross et al. *Biochem. Biophys. Res. Commun.* 302 354-358 (2003).

[0260] A. Connective Tissue Growth Factor

Real time PCR analysis		
Comparison	Expression Fold change	t test
Diabetics + Tro (10)/Diabetics - Tro (10)	0.35	0.042

Legend "Tro" indicates treatment with Troglitazone. "Fold change" indicates fold change in CTGF expression calculated as the ratio of mean diabetic post-Tro expression/mean diabetic pre-Tro expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0261] These data further show that treatment with Troglitazone reduces the expression of CTGF in muscle from diabetic individuals.

Glut 4 translocation analysis		
Insulin (nM)	Fold Change (Mean hCTGF/Mean LacZ) (n = 3)	t test (hCTGF vs LacZ)
0	0.97	0.960
1	0.35	0.044
10	0.45	0.010

Legend "Fold Change" indicates the following ratio; (Mean % of CTGF-expressing cells scored positive for cell surface Glut 4)/(Mean % of LacZ-expressing cells scored positive for cell surface Glut 4). "n" is the number of replicates.

[0262] These data show that increasing the levels of hCTGF in 3T3-L1 adipocytes significantly inhibited Glut 4 translocation at all insulin concentrations tested with little or no effect on basal Glut 4 translocation.

Glucose transport analysis in 3T3-L1 adipocytes							
Insulin (nM)	Expt 1		Expt 2		Expt 3		Mean FC (hCTGF/Con) +/- SEM (n = 3)
	FC (hCTGF/Con)	t test	FC (hCTGF/Con)	t test	FC (hCTGF/Con)	t test	
0	0.7	0.056	0.6	0.085	0.8	0.116	0.70 +/- 0.03
0.05	0.6	0.076	0.6	2.11E-04	0.7	0.233	0.63 +/- 0.06
0.1	0.6	0.038	0.7	0.120	0.7	0.183	0.67 +/- 0.02
0.3	0.7	0.057	0.8	0.130	0.8	0.395	0.76 +/- 0.05
1	0.8	0.027	0.8	0.180	0.9	0.308	0.83 +/- 0.02
10	0.6	0.015	0.7	0.012	0.9	0.143	0.75 +/- 0.10

Legend "Con" indicates control cells which do not express hCTGF. "FC" indicates the fold change defined as the following ratio; glucose transport in hCTGF-expressing cells/glucose transport in non-hCTGF-expressing cells. h" is human. "n" is the number of replicates. SEM is the standard error of the mean.

[0263] These data show that increasing the level of nCTGF in 3T3-L1 adipocytes inhibited both basal and insulin stimulated glucose transport.

Glucose transport analysis in L6 myotubes							
Insulin (nM)	Expt 1		Expt 2		Expt 3		Mean FC (CTGF/Con) +/- SEM (n = 3)
	FC (hCTGF/Con)	t test	FC (hCTGF/Con)	t test	FC (hCTGF/Con)	t test	
0	0.78	2.49E-04	0.79	0.042	0.79	3.6E-05	0.79 +/- 0.004
10	0.67	0.163	0.92	0.121	0.77	0.080	0.79 +/- 0.07
100	0.70	0.226	1.05	0.404	0.74	0.001	0.83 +/- 0.11

Legend "Con" indicates control cells which do not express CTGF. "FC" indicates the fold change defined as the following ratio; glucose transport in CTGF-expressing cells/glucose transport in non-CTGF-expressing cells. "h" is human. "n" is the number of replicates. SEM is the standard error of the mean.

[0264] These data show that increasing the levels of hCTGF in L6 myotubes inhibited both basal and insulin stimulated glucose transport. These data are consistent with the data obtained in 3T3-L1 adipocytes.

[0265] Decreased CTGF levels such as those observed in diabetic patients treated with troglitazone would therefore be beneficial at improving glucose uptake into muscle and fat.

[0266] B. TIEG

Real time PCR analysis		
Comparison	Expression Fold change	t test
Diabetics + Tro (9)/Diabetics - Tro (9)	0.69	0.057
Diabetics (19)/Lean (17)	2.56	0.005

Legend "Tro" indicates treatment with Troglitazone. "Fold change" indicates fold change in TIEG expression calculated as either the ratio of mean diabetic post-Tro expression/mean diabetic pre-Tro expression or mean diabetic pre-Tro expression/mean lean pre-Tro expression. Numbers in parentheses indicate the number of patient samples analyzed by real time PCR.

[0267] These data further show that TIEG is significantly over-expressed in muscle from diabetic individuals when

compared to muscle from lean individuals. Furthermore the data also shows that treatment with Troglitazone reduces the expression of TEG in muscle from diabetic individuals.

Glut 4 translocation analysis		
Insulin (nM)	Fold Change (Mean hTIEG/Mean LacZ) (n = 3)	t test (hTIEG vs LacZ)
0	1.34	0.600
1	0.20	0.019
10	0.11	5.1E-05

Legend "Fold Change" indicates the following ratio; (Mean % of TIEG-expressing cells scored positive for cell surface Glut 4)/(Mean % of LacZ-expressing cells scored positive for cell surface Glut 4). "h" is human. "n" is the number of replicates.

[0268] These data show that increasing the levels of hTIEBG in 3T3-L1 adipocytes significantly inhibited Glut 4 translocation at all insulin concentrations tested with little or no effect on basal Glut 4 translocation. Increased TIEG levels such as those observed in human diabetic muscle are therefore inhibitory to insulin action.

SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 16

<210> SEQ ID NO 1
<211> LENGTH: 2312
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human connective tissue growth factor (CTGF)
      cDNA
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (146)..(1195)
<223> OTHER INFORMATION: CTGF
    
```

-continued

<400> SEQUENCE: 1

```

tccagtgcg gagccgccc gccgacagc ccgagacgac agcccggcgc gtcccggctc   60
ccacctcoga ccaccgccag cgctccaggc cccgcgctcc ccgctcgcgc ccaccgcgcc   120
ctccgctccg ccgcagtgcc caaccatgac cgcgcgagc atggggcccgc tccgctcgc   180
cttcgtggtc ctccctgccc tctgcagccg gccggccgct gcccagaact gcagcgggcc   240
gtgcgctgtc ccggacgagc cggcgccgcg ctgcccggcg gccgtgagcc tcgtgctgga   300
cggctgcggc tgtgcccgcg tctgcgcca gcagctgggc gagctgtgca ccgagcgcga   360
cccctgcgac ccgcacaagc gcctctctg tgacttcggc tcccggcca accgcaagat   420
cggcgtgtgc accgcaaaag atggtgctcc ctgcatcttc ggtggtacgg tgtaccgcag   480
cggagagtcc ttccagagca gctgcaagta ccagtgcacg tgcctggacg gggcggtggg   540
ctgcatgccc ctgtgcagca tggacgttcg tctgcccagc cctgactgcc cttcccagag   600
gaggggtcaag ctgcccggga aatgctgcga ggagtgggtg tgtgacgagc ccaaggacca   660
aacctgggtt gggcctgccc tcgcggctta ccgactggaa gacacgtttg gccagacccc   720
aactatgatt agagccaact gcctgttcca gaccacagag tggagcgcct gttccaagac   780
ctgtgggatg gccatctcca cccgggttac caatgacaac gcctcctgca ggctagagaa   840
gcagagccgc ctgtgcatgg tcaggccttg cgaagctgac ctggaagaga acattaagaa   900
gggcaaaaag tgcacccgta ctccccaaat ctccaagcct atcaagtttg agctttctgg   960
ctgcaccagc atgaagacat accgagctaa attctgtgga gtatgtaccg acggccgatg  1020
ctgcaccccc cacagaacca ccaccctgcc ggtggagttc aagtgcctg acggcgaggt  1080
catgaagaag aacatgatgt tcatcaagac ctgtgcctgc cattacaact gtcccggaga  1140
caatgacatc tttgaatcgc tgtactacag gaagatgtac ggagacatgg catgaagcca  1200
gagagtgaga gacattaact cattagactg gaacttgaac tgattccat ctcatTTTTc  1260
cgtaaaaatg atttcagtag cacaagttat ttaaactctgt ttttctaact gggggaaaag  1320
attccccccc aattcaaaac attgtgcoat gtcaaaaaa tagtctatct tcccagaca  1380
ctggtttgaa gaatgttaag acttgacagt ggaactacat tagtacacag caccagaatg  1440
tatattaagg tgtggcttta ggagcagtg gaggtacca gcagaaagg tagtatcatc  1500
agatagctct tatacagtag atatgcctgc ttttgaagt gtaattgaga aggaaaattt  1560
tagcgtgctc actgacctgc ctgtagcccc agtgacagct aggatgtgca ttctccagcc  1620
atcaagagac tgagtcaagt tgttccttaa gtcagaacag cagactcagc tctgacattc  1680
tgattcgaat gacactgttc aggaatcggg atcctgtcga ttagactgga cagcttgtgg  1740
caagtgaatt tcctgtaaca agccagattt tttaaaattt atattgtaa tttgtgtgt  1800
gtgtgtgtgt gtgtatatat atatatatat gtacagttat ctaagttaat ttaaagttgt  1860
ttgtgccttt ttatTTTTgt ttttaatgct ttgatatTtc aatgttagcc tcaatttctg  1920
aacaccatag gtagaatgta aagcttctct gatcgttcaa agcatgaaat ggatacttat  1980
atggaattc tctcagatag aatgacagtc cgtcaaaaca gattgtttgc aaaggggagg  2040
catcagtgtc cttggcaggc tgatttctag gtaggaaatg tggtagctca cgtcaccttt  2100
taatgaacaa atggccttta ttaaaaactg agtgactcta tatagctgat cagttttttc  2160
acctggaagc atttgtttct actttgatat gactgttttt cggacagttt atttgttgag  2220

```

-continued

 agtgtgacca aaagttacat gtttgacacct ttctagtgtga aaataaagta tattttttct 2280

aaaaaaaaaa aaaaacgaca gcaacggaat tc 2312

<210> SEQ ID NO 2

<211> LENGTH: 349

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<220> FEATURE:

<223> OTHER INFORMATION: human connective tissue growth factor (CTGF)

<400> SEQUENCE: 2

 Met Thr Ala Ala Ser Met Gly Pro Val Arg Val Ala Phe Val Val Leu
 1 5 10 15

 Leu Ala Leu Cys Ser Arg Pro Ala Val Gly Gln Asn Cys Ser Gly Pro
 20 25 30

 Cys Arg Cys Pro Asp Glu Pro Ala Pro Arg Cys Pro Ala Gly Val Ser
 35 40 45

 Leu Val Leu Asp Gly Cys Gly Cys Cys Arg Val Cys Ala Lys Gln Leu
 50 55 60

 Gly Glu Leu Cys Thr Glu Arg Asp Pro Cys Asp Pro His Lys Gly Leu
 65 70 75 80

 Phe Cys Asp Phe Gly Ser Pro Ala Asn Arg Lys Ile Gly Val Cys Thr
 85 90 95

 Ala Lys Asp Gly Ala Pro Cys Ile Phe Gly Gly Thr Val Tyr Arg Ser
 100 105 110

 Gly Glu Ser Phe Gln Ser Ser Cys Lys Tyr Gln Cys Thr Cys Leu Asp
 115 120 125

 Gly Ala Val Gly Cys Met Pro Leu Cys Ser Met Asp Val Arg Leu Pro
 130 135 140

 Ser Pro Asp Cys Pro Phe Pro Arg Arg Val Lys Leu Pro Gly Lys Cys
 145 150 155 160

 Cys Glu Glu Trp Val Cys Asp Glu Pro Lys Asp Gln Thr Val Val Gly
 165 170 175

 Pro Ala Leu Ala Ala Tyr Arg Leu Glu Asp Thr Phe Gly Pro Asp Pro
 180 185 190

 Thr Met Ile Arg Ala Asn Cys Leu Val Gln Thr Thr Glu Trp Ser Ala
 195 200 205

 Cys Ser Lys Thr Cys Gly Met Gly Ile Ser Thr Arg Val Thr Asn Asp
 210 215 220

 Asn Ala Ser Cys Arg Leu Glu Lys Gln Ser Arg Leu Cys Met Val Arg
 225 230 235 240

 Pro Cys Glu Ala Asp Leu Glu Glu Asn Ile Lys Lys Gly Lys Lys Cys
 245 250 255

 Ile Arg Thr Pro Lys Ile Ser Lys Pro Ile Lys Phe Glu Leu Ser Gly
 260 265 270

 Cys Thr Ser Met Lys Thr Tyr Arg Ala Lys Phe Cys Gly Val Cys Thr
 275 280 285

 Asp Gly Arg Cys Cys Thr Pro His Arg Thr Thr Thr Leu Pro Val Glu
 290 295 300

 Phe Lys Cys Pro Asp Gly Glu Val Met Lys Lys Asn Met Met Phe Ile
 305 310 315 320

Lys Thr Cys Ala Cys His Tyr Asn Cys Pro Gly Asp Asn Asp Ile Phe

-continued

325	330	335	
Glu Ser Leu Tyr Tyr Arg Lys Met Tyr Gly Asp Met Ala			
340	345		
<p><210> SEQ ID NO 3 <211> LENGTH: 2330 <212> TYPE: DNA <213> ORGANISM: Mus musculus <220> FEATURE: <223> OTHER INFORMATION: mouse connective tissue growth factor (CTGF) homolog cDNA <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (204)..(1250) <223> OTHER INFORMATION: CTGF homolog</p>			
<p><400> SEQUENCE: 3</p>			
agactcagcc agatccactc cagctccgac cccagagagac cgacctcctc cagacggcag			60
cagccccagc ccagccgcaca accccagacg ccaccgcctg gagcgtccag acaccaacct			120
ccgccccctgt ccgaatccag gctccagccg cgcctctcgt cgcctctgca cctgctgtg			180
catcctccta ccgctgcccg atcatgctcg cctccgtcgc aggtcccatc agcctcgcct			240
tgggtctcct cgcctctcgc acccggcctg ctacgggcca ggactgcagc gcgcaatgtc			300
agtgcgcagc cgaagcagcg ccgcactgcc ccgccggcgt gagcctggtg ctggacggct			360
gcggctgctg ccgctctcgc gccaaagcagc tgggagaact gtgtacggag cgtgaccct			420
gcgaccaca caagggcctc ttctgcgatt tcggctcccc cgccaaccgc aagattggag			480
tgtgactcgc caaagatggg gcaccctgtg tcttcgggtg gtcggtgtac cgcagcggtg			540
agtccttcca aagcagctgc aaataccaat gcacttgctt ggatggggcc gtgggctcgc			600
tgcccctatg cagcatggac gtgcgcctgc ccagccctga ctgccccttc ccgagaaggg			660
tcaagctgcc tgggaaatgc tgcgaggagt ggggtgtgta cgagcccagg gaccgcacag			720
cagttggccc tgcccctagct gcctaccgac tggaaagacac atttggccca gaccacaacta			780
tgatgcgagc caactgcctg gtccagacca cagagtggag cgcctgttct aagacctgtg			840
gaatgggcat ctccaccga gttaccaatg acaatacctt ctgcagactg gagaagcaga			900
gccgcctctg catggtcagg ccctgcgaag ctgacctgga ggaaaacatt aagaagggca			960
aaaagtgcac ccggacacct aaaatcgcca agcctgtcaa gtttgagctt tctggctgca			1020
ccagtgtgaa gacatacagc gctaagtctt gcggggtgtg cacagacggc cgctgctgca			1080
caccgcacag aaccaccact ctgccagtg agttcaaatg ccccgatggc gagatcatga			1140
aaaagaatat gatgttcac cagacctgtg cctgccatta caactgtcct ggggacaatg			1200
acatctttga gtcctgtac tacaggaaga tgtacggaga catggcgtaa agccaggaag			1260
taaggacac gaactcatta gactataact tgaactgagt tgcattctcat tttctctgt			1320
aaaaacaatt acagtgcac attaatata atctgtgttt ttaactaccg tgggaggaac			1380
tatccccaca aagtgagaac gttatgtcat ggccatacaa gtagtctgtc aacctcagac			1440
actggtttcg agacagtta cacttgacag ttgttcatta gcgcacagtg ccagaacgca			1500
cactgaggtg agtctcctgg aacagtggag atgccaggag aaagaaagac aggtactagc			1560
tgaggttatt ttaaaagcag cagtgtgctt actttttgga gtgtaaccgg ggagggaaat			1620
tatagcatgc ttgcagacag acctgtccta gcgagagctg agcatgtgtc ctccactaga			1680

-continued

```

tgaggctgag tccagctggt cttaagaac agcagtttca gctctgacca ttctgattcc 1740
agtgcacatt gtcaggagtc agagccttgt ctgtagact ggacagcttg tggcaagtaa 1800
gtttgctgt aacaagccag atttttattg atattgtaaa tattgtggat atatatatat 1860
atatatatat atttgtacag ttatctaagt taatttaaag tcatttggtt ttgttttaag 1920
tgcttttggg attttaaact gatagcctca aactccaaac accataghta ggacacgaag 1980
cttatctgtg attcaaaaca aaggagatac tgcagtggga attgtgacct gactgactct 2040
ctgtcagaac aaacaaatgc tgtgcagggt ataaagctat gtattggaag tcagatttct 2100
agtaggaaat gtggtaaat ccctgttggg gaacaaatgg cctttattaa gaaatggctg 2160
gctcagggta aggtccgatt cctaccagga agtgcttgct gottotttga ttatgactgg 2220
tttgggtg ggggcagttt atttgttgg agtgtgacca aaagtacat gtttgcactt 2280
tctagttgaa aataaagtat atatatattt ttatatgaaa aaaaaaaaaa 2330

```

```

<210> SEQ ID NO 4
<211> LENGTH: 348
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse connective tissue growth factor (CTGF)
homolog

```

<400> SEQUENCE: 4

```

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

```

-continued

225		230		235		240									
Cys	Glu	Ala	Asp	Leu	Glu	Glu	Asn	Ile	Lys	Lys	Gly	Lys	Lys	Cys	Ile
				245					250					255	
Arg	Thr	Pro	Lys	Ile	Ala	Lys	Pro	Val	Lys	Phe	Glu	Leu	Ser	Gly	Cys
			260					265						270	
Thr	Ser	Val	Lys	Thr	Tyr	Arg	Ala	Lys	Phe	Cys	Gly	Val	Cys	Thr	Asp
			275				280						285		
Gly	Arg	Cys	Cys	Thr	Pro	His	Arg	Thr	Thr	Thr	Leu	Pro	Val	Glu	Phe
	290					295					300				
Lys	Cys	Pro	Asp	Gly	Glu	Ile	Met	Lys	Lys	Asn	Met	Met	Phe	Ile	Lys
	305				310					315					320
Thr	Cys	Ala	Cys	His	Tyr	Asn	Cys	Pro	Gly	Asp	Asn	Asp	Ile	Phe	Glu
				325					330					335	
Ser	Leu	Tyr	Tyr	Arg	Lys	Met	Tyr	Gly	Asp	Met	Ala				
			340					345							

<210> SEQ ID NO 5
 <211> LENGTH: 2345
 <212> TYPE: DNA
 <213> ORGANISM: Rattus norvegicus
 <220> FEATURE:
 <223> OTHER INFORMATION: rat connective tissue growth factor (CTGF)
 homolog cDNA
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (225)..(1268)
 <223> OTHER INFORMATION: CTGF homolog

<400> SEQUENCE: 5

```

cacagctctt ctctccaaga agactcagcc agaccactc cagctccgac cctaggagac    60
cgacctcctc cagacggcag cagccccagc ccagtgagaca accccaggag ccaccacctg    120
gagcgtccgg acaccaacct ccgccccgag accgagtgcca ggctccggcc gcgcccctcg    180
tcgcctctgc accccgctgt gcgtctctct gccgcgcccc gaccatgctc gcctccgctg    240
cgggtccctg tagcctcgcc ttggtgctcc tcctctgcac ccggcctgcc accggccagg    300
actgcagcgc gcagtgctag tgccgagctg aagcggcgcc gcgctgcccc gccggcgtga    360
gcctgggtgct ggacggctgc ggctgctgcc gcgtctgcgc caagcagctg ggagaactgt    420
gcacggagcg tgatccctgc gaccacaca agggctctct ctgcgacttc ggctcccccg    480
ccaaccgcaa gattggcgtg tgcaactgcca aagatggctg accctgtgct tcggtgggt    540
ccgtgtaccg cagcggcgag tccttccaaa gcagttgcaa ataccagtgc acttgccctg    600
atggggccgt gggctgtgtg cccctgtgca gcatggacgt ggcctgccc agccctgact    660
gcccttccc  gagaagggtc aagctgccc  ggaaatgctg tgaggagtgg gtgtgtgatg    720
agcccaagga ccgcacagtg gttggccctg ccctagctgc ctaccgactg gaagacacat    780
ttggccctga cccaactatg atgcgagcca actgcctggt ccagaccaca gactggagcg    840
cctgttctaa gacctgtggg atgggcatct ccaccgggt  taccaatgac aataccttct    900
gcaggctgga gaagcagagt cgtctctgca tggtcaggcc ctgtgaagct gacctagagg    960
aaaacattaa gaagggcaaa aagtgcattc ggacgctaa aattgccaag cctgtcaagt   1020
ttgagctttc tggctgcacc agtgtgaaga cctaccgggc taagttctgt ggggtgtgca   1080
cggacggccc ctgctgcaca ccgcacagaa ccaccacact gccggtggag ttcaagtgcc   1140
    
```

-continued

```

ccgatggcga gatcatgaaa aagaacatga tgttcatcaa gacctgtgcc tgccattaca 1200
actgtcccgg ggacaatgac atctttgagt ccttgtacta caggaagatg tatggagaca 1260
tggcgtaaag ccagggagta agggacacga actcatttag actataactt gaactgagtt 1320
acatctcatt ttcttctgta aaaaaacaaa aaggattaca gtagcacatt aatttaaadc 1380
tgggttccta actgctgtgg gagaaaaacac cccaccgaag tgagaaccgt gtgtcattgt 1440
catgcaaata gcctgtcaat ctcagacact ggtttcgaga cagtttagac ttgacagttg 1500
ttcactagcg cacagtgaca gaacgcacac taagtgagc ctcctggaag agtgagatg 1560
ccaggagaaa gacaggact agctgaggtc attttaaaag cagcgatag cctacttttt 1620
ggagtgtgac aggggaggga cattatagct tgcttgaca cagacctgct ctgcaagag 1680
ctgggtgtgt gtcctccact cggtgaggct gaagccagct attctttcag taagaacagc 1740
agtttcagcg ctgacattct gattccagtg aactggctg ggagtcagaa ccttgtctat 1800
tagactggac agcttgggc aagtgaatth gccgtaaca agccagatth ttatggatct 1860
tgtaaatatt gtggataaat atatataatth gtacagttat ctaagttaat ttaaagacgt 1920
ttgtgcctat tgttcttght ttaagtgcct ttggaattht taaactgata gcctcaaac 1980
ccaaacacca tcgataggac ataaagcttg tctgtgattc aaaacaaag agatactgca 2040
gtggaactg taacctgagt gactgtctgt cagaacatat ggtacgtaga cggtaaagca 2100
atggatcaga agtcagatth ctatagggaa atgtaaaatc actgttgccg aacaaatggc 2160
ctttattaag aaatggcttg ctcagggtaa ctggtcagat ttccacgagg aagtgtttgc 2220
tgcttctttg actatgactg gtttgggagg cagtttatth gttgagatg tgacaaaag 2280
ttacatgtht gcacctthct agttgaaat aaagtatata tttttata aaaaaaaaaa 2340
aaaaa 2345

```

```

<210> SEQ ID NO 6
<211> LENGTH: 347
<212> TYPE: PRT
<213> ORGANISM: Rattus norvegicus
<220> FEATURE:
<223> OTHER INFORMATION: rat connective tissue growth factor (CTGF)
homolog

```

<400> SEQUENCE: 6

```

Met Leu Ala Ser Val Ala Gly Pro Val Ser Leu Ala Leu Val Leu Leu
 1             5             10            15
Leu Cys Thr Arg Pro Ala Thr Gly Gln Asp Cys Ser Ala Gln Cys Gln
 20            25            30
Cys Ala Ala Glu Ala Ala Pro Arg Cys Pro Ala Gly Val Ser Leu Val
 35            40            45
Leu Asp Gly Cys Gly Cys Cys Arg Val Cys Ala Lys Gln Leu Gly Glu
 50            55            60
Leu Cys Thr Glu Arg Asp Pro Cys Asp Pro His Lys Gly Leu Phe Cys
 65            70            75            80
Asp Phe Gly Ser Pro Ala Asn Arg Lys Ile Gly Val Cys Thr Ala Lys
 85            90            95
Asp Gly Ala Pro Cys Val Phe Gly Gly Ser Val Tyr Arg Ser Gly Glu
100           105           110
Ser Phe Gln Ser Ser Cys Lys Tyr Gln Cys Thr Cys Leu Asp Gly Ala
115           120           125

```

-continued

Val Gly Cys Val Pro Leu Cys Ser Met Asp Val Arg Leu Pro Ser Pro
 130 135 140
 Asp Cys Pro Phe Pro Arg Arg Val Lys Leu Pro Gly Lys Cys Cys Glu
 145 150 155 160
 Glu Trp Val Cys Asp Glu Pro Lys Asp Arg Thr Val Val Gly Pro Ala
 165 170 175
 Leu Ala Ala Tyr Arg Leu Glu Asp Thr Phe Gly Pro Asp Pro Thr Met
 180 185 190
 Met Arg Ala Asn Cys Leu Val Gln Thr Thr Glu Trp Ser Ala Cys Ser
 195 200 205
 Lys Thr Cys Gly Met Gly Ile Ser Thr Arg Val Thr Asn Asp Asn Thr
 210 215 220
 Phe Cys Arg Leu Glu Lys Gln Ser Arg Leu Cys Met Val Arg Pro Cys
 225 230 235 240
 Glu Ala Asp Leu Glu Glu Asn Ile Lys Lys Gly Lys Lys Cys Ile Arg
 245 250 255
 Thr Pro Lys Ile Ala Lys Pro Val Lys Phe Glu Leu Ser Gly Cys Thr
 260 265 270
 Ser Val Lys Thr Tyr Arg Ala Lys Phe Cys Gly Val Cys Thr Asp Gly
 275 280 285
 Arg Cys Cys Thr Pro His Arg Thr Thr Thr Leu Pro Val Glu Phe Lys
 290 295 300
 Cys Pro Asp Gly Glu Ile Met Lys Lys Asn Met Met Phe Ile Lys Thr
 305 310 315 320
 Cys Ala Cys His Tyr Asn Cys Pro Gly Asp Asn Asp Ile Phe Glu Ser
 325 330 335
 Leu Tyr Tyr Arg Lys Met Tyr Gly Asp Met Ala
 340 345

<210> SEQ ID NO 7
 <211> LENGTH: 3085
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens
 <220> FEATURE:
 <223> OTHER INFORMATION: human transforming growth factor-beta (TGFB)
 inducible early growth response (TIEG) cDNA
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (320)..(1729)
 <223> OTHER INFORMATION: TIEG
 <400> SEQUENCE: 7

ctcacgctct ctttcctgc cgcctgcctt tcttttttcc tttttttgta ttggcgtctt 60
 ggggctgta cacacacgcg cgctgtccat tgcagcttac ataaaggcgg gcgcgattat 120
 gcaattatat tgttagcgat atttcaagag caatggctcg ttttcttagg atttcaacac 180
 gaaggcatca tgcatttttg aaaaactagt attgagaata ataccttga acgtaaagaa 240
 tgttttttgg tatttttaca caatctctac ttgacccaaa cgagtctgga cagttttctt 300
 ttaatgaaaa ataggagaaa tggaggaaa aatggaaatg atttctgaaa ggccaaaaga 360
 gagtatgtat tcttggaaa aaactgcaga gaaaagtgat tttgaagctg tagaagcact 420
 tatgtcaatg agctgcagtt ggaagtctga ttttaagaaa tacgttgaaa acagacctgt 480
 tacaccagta tctgatttgt cagaggaaga gaatctgctt ccgggaacac ctgattttca 540

-continued

tacaatccca	gcattttggt	tgactccacc	ttacagtcct	tctgactttg	aaccctctca	600
agtgtcaaat	ctgatggcac	cagcgccatc	tactgtacac	ttcaagtcac	tctcagatac	660
tgccaaacct	cacattgccg	cacctttcaa	agaggaagaa	aagagcccag	tatctgcccc	720
caaaactcccc	aaagctcagg	caacaagtgt	gattcgtcat	acagctgatg	cccagctatg	780
taaccaccag	acctgcccac	tgaagcagc	cagcatcctc	aactatcaga	acaattcctt	840
tagaagaaga	accaccta	atgttgaggc	tgcaagaaag	aacataccat	gtgccgctgt	900
gtcaccaaac	agatccaaat	gtgagagaaa	cacagtggca	gatgttgatg	agaaagcaag	960
tgctgcaact	tatgactttt	ctgtgccttc	ctcagagacg	gtcatctgca	ggtctcagcc	1020
agcccctgtg	tccccacaac	agaagtcagt	gttggtctct	ccacctgcag	tatctgcagg	1080
gggagtgcc	cctatgccg	tcacttgcca	gatggttccc	cttctgcca	acaacctgt	1140
tgtgacaaca	gtcgttccca	gcaactctcc	cagccagcca	ccagccgttt	gccccctgt	1200
tgtgttcatt	ggcacacaag	tccccaaagg	cgctgtcatg	tttgtgttac	cccagcccgt	1260
tgtgcagagt	tcaaaagctc	cggtgggtgag	cccgaatggc	accagactct	ctcccattgc	1320
ccctgctcct	gggttttccc	cttcagcagc	aaaagtcact	cctcagattg	attcatcaag	1380
gataaggagt	cacatctgta	gccaccagc	atgtggcaag	acatacttta	aaagttccca	1440
tctgaaggcc	cacacgagga	cgcacacagc	agaaaagcct	ttcagctgta	gctggaaagg	1500
ttgtgaaagg	aggtttgccc	gttctgatga	actgtccaga	cacaggcga	cccacacggg	1560
tgagaagaaa	tttgcgtgcc	ccatgtgtga	ccggcggttc	atgaggagtg	accatttgac	1620
caagcatgcc	cggcgccatc	tatcagccaa	gaagctacca	aactggcaga	tggaagtgag	1680
caagctaaat	gacattgctc	tacctccaac	ccctgctccc	acacagtgac	agaccggaaa	1740
gtgaagagtc	agaactaaat	ttggtctcag	cgggagccag	tggtgatgta	aaaatgcttc	1800
cactgcaagt	ctgtggcccc	acaacgtggg	cttaaagcag	aagccccaca	gcctggcacg	1860
aaggccccgt	ctgggttagg	tgactaaaag	ggcttcggcc	acaggcaggt	cacagaaagg	1920
caggtttcat	ttcttatc	ataagagaga	tgagaagcct	tttattcctt	tgaatatatt	1980
ttgaagggtt	cagatgaggt	caacacaggt	agcacagatt	ttgaaactgt	gtgcatattt	2040
gttactttac	ttttgtggtt	tatacttgag	accaactttt	caatgtgatt	cttctaaagc	2100
actggtttca	agaatatgga	ggctggaagg	aaataaacat	tacggtacag	acatggagat	2160
gtaaatgag	tttgtattat	tacaaatatt	gtcatctttt	tctagagtta	tcttctttat	2220
tattcctagt	ctttccagtc	aacatcgtgg	atgtagtgat	taaatatatc	tagaactatc	2280
atttttacac	tattgtgaat	atttggaatt	gaacgactgt	atattgctaa	gagggcccaa	2340
agaattggaa	tcctccttaa	tttaattgct	ttgaagcata	gctacaattt	gtttttgcat	2400
ttttgttttg	aaagtttaac	aatgactgt	atctaggcat	ttcattatgc	tttgaacttt	2460
agtttgccctg	cagtttcttg	tgtagatttg	aaaattgtat	accaatgtgt	tttctgtaga	2520
ctctaagata	cactgcactt	tgtttagaaa	aaaaactgaa	gatgaaatat	atattgtaaa	2580
gaagggat	taagaatcct	agataaactc	ttgaaaaaga	tggttatgt	catcagtaaa	2640
gtacctttat	ggtatgagga	tataatgtgt	gctttattga	attagaaaat	tagtgaccat	2700
tattcacagg	tgacaaaatg	ttgtcctggt	aatttatagg	agttttttgg	ggatgtggag	2760
gtagttgggt	agaaaaatta	ttagaacatt	cacttttgtt	aacagtat	ctcttttatt	2820

-continued

```

ctggtatata gtggatgata tacacagtgg caaaacaaaa gtacattgct taaaatata 2880
agtgaaaaat gtcactatata cttccattt aacattgttt ttgtatattg ggtgtagatt 2940
tctgacatca aaacttgac ccttgaaaa caaaagtttt aattaaaaaa aatccttgtg 3000
acttacaatt tgcacaatat ttcttttgtt gtactttata tcttgtttac aataaagaat 3060
tccctttggt aaaaaaaaa aaaaa 3085

```

```

<210> SEQ ID NO 8
<211> LENGTH: 469
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human transforming growth factor-beta (TGFB)
        inducible early growth response (TIEG)

```

```

<400> SEQUENCE: 8

```

```

Met Glu Glu Arg Met Glu Met Ile Ser Glu Arg Pro Lys Glu Ser Met
  1           5           10           15
Tyr Ser Trp Asn Lys Thr Ala Glu Lys Ser Asp Phe Glu Ala Val Glu
          20           25           30
Ala Leu Met Ser Met Ser Cys Ser Trp Lys Ser Asp Phe Lys Lys Tyr
          35           40           45
Val Glu Asn Arg Pro Val Thr Pro Val Ser Asp Leu Ser Glu Glu Glu
          50           55           60
Asn Leu Leu Pro Gly Thr Pro Asp Phe His Thr Ile Pro Ala Phe Cys
          65           70           75           80
Leu Thr Pro Pro Tyr Ser Pro Ser Asp Phe Glu Pro Ser Gln Val Ser
          85           90           95
Asn Leu Met Ala Pro Ala Pro Ser Thr Val His Phe Lys Ser Leu Ser
          100          105          110
Asp Thr Ala Lys Pro His Ile Ala Ala Pro Phe Lys Glu Glu Lys
          115          120          125
Ser Pro Val Ser Ala Pro Lys Leu Pro Lys Ala Gln Ala Thr Ser Val
          130          135          140
Ile Arg His Thr Ala Asp Ala Gln Leu Cys Asn His Gln Thr Cys Pro
          145          150          155          160
Met Lys Ala Ala Ser Ile Leu Asn Tyr Gln Asn Asn Ser Phe Arg Arg
          165          170          175
Arg Thr His Leu Asn Val Glu Ala Ala Arg Lys Asn Ile Pro Cys Ala
          180          185          190
Ala Val Ser Pro Asn Arg Ser Lys Cys Glu Arg Asn Thr Val Ala Asp
          195          200          205
Val Asp Glu Lys Ala Ser Ala Ala Leu Tyr Asp Phe Ser Val Pro Ser
          210          215          220
Ser Glu Thr Val Ile Cys Arg Ser Gln Pro Ala Pro Val Ser Pro Gln
          225          230          235          240
Gln Lys Ser Val Leu Val Ser Pro Pro Ala Val Ser Ala Gly Gly Val
          245          250          255
Pro Pro Met Pro Val Ile Cys Gln Met Val Pro Leu Pro Ala Asn Asn
          260          265          270
Pro Val Val Thr Thr Val Val Pro Ser Thr Pro Pro Ser Gln Pro Pro
          275          280          285
Ala Val Cys Pro Pro Val Val Phe Met Gly Thr Gln Val Pro Lys Gly

```

-continued

290	295	300
Ala Val Met Phe Val Val Pro Gln Pro Val Val Gln Ser Ser Lys Pro		
305	310	315 320
Pro Val Val Ser Pro Asn Gly Thr Arg Leu Ser Pro Ile Ala Pro Ala		
	325	330 335
Pro Gly Phe Ser Pro Ser Ala Ala Lys Val Thr Pro Gln Ile Asp Ser		
	340	345 350
Ser Arg Ile Arg Ser His Ile Cys Ser His Pro Gly Cys Gly Lys Thr		
	355	360 365
Tyr Phe Lys Ser Ser His Leu Lys Ala His Thr Arg Thr His Thr Gly		
	370	375 380
Glu Lys Pro Phe Ser Cys Ser Trp Lys Gly Cys Glu Arg Arg Phe Ala		
385	390	395 400
Arg Ser Asp Glu Leu Ser Arg His Arg Arg Thr His Thr Gly Glu Lys		
	405	410 415
Lys Phe Ala Cys Pro Met Cys Asp Arg Arg Phe Met Arg Ser Asp His		
	420	425 430
Leu Thr Lys His Ala Arg Arg His Leu Ser Ala Lys Lys Leu Pro Asn		
	435	440 445
Trp Gln Met Glu Val Ser Lys Leu Asn Asp Ile Ala Leu Pro Pro Thr		
	450	455 460
Pro Ala Pro Thr Gln		
465		

<210> SEQ ID NO 9
 <211> LENGTH: 3045
 <212> TYPE: DNA
 <213> ORGANISM: Mus musculus
 <220> FEATURE:
 <223> OTHER INFORMATION: mouse transforming growth factor-beta (TGFB)
 inducible early growth response (TIEG) homolog
 cDNA
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (114)..(1553)
 <223> OTHER INFORMATION: TIEG homolog
 <400> SEQUENCE: 9

```

cagacagcgc ggagcgcggt ggcgtcgacg tctagtgtct cagtgtctccc gtctgtggct    60
aactaagcag ccagcagcca ggcagctcgc gacctgcggc caggcagcca accatgctca    120
acttcggcgc ttctctccag caagcttcgg aggggaaaat ggaactaatt tctgaaaagc    180
ccagagaggg gatgcatccc tgggacaaaag ctgagcagag tgactttgaa gcggtggaag    240
cgctcatgtc catgagctgc gactggaagt ctcatttcaa gaaatacctt gaaaacaggc    300
ctgtcacacc agtgtctgat acctccgagg atgacagctt gcttccaggg acgcctgacc    360
ttcagacagt cccagcattt tgtttaacgc caccttacag cccctctgac ttcgaaccct    420
cccaagggtc aaatctgact gcatcagcgc catctactgg ccaactcaaa tctttctccg    480
atgttgccaa gcctccaggc gccactcctt tcaaagagga ggaaaagaat ccttttagctg    540
cccctctctt tcctaaggct caagccacca gtgtcatccg tcacacagct gatgcccaac    600
tgtgcaacca ccagtcctgc cccgtgaaag cagctagcat cctcaactat caggacaatt    660
ctttccggag aagaaccac ggaaatgttg aggtactctg aaagaacata cctgtgtctg    720
cagtgtcacc aaacagatcc aagcctgagc ccagcacagt gtccgatggt gatgagaagg    780
    
```

-continued

cgggcgctgc actatatgac tttgctgtgc cttcctcaga gacagtaatt tgtaggtctc	840
agccagctcc ttcgtcccca gtgcagaagt cagtactggt gtcttcacct acagtatcca	900
ctgggggagt gccaccctcg cctgtcatct gccagatggt tccccttctt gccacaact	960
ctcttgttag cacagtgtc cccagcactc ctcttagcca gccaccagct gtctgtctac	1020
ctgtgttgtt catgggcact caggtgcctg agggcaccgt cgtgtttgtg gtaccccagc	1080
ccgttgtgca gagcccaagg cctccagtg tgagcccag tggcaccaga ctgtctcca	1140
ttgccctgc tcttgatct tctcctcag cagcaagggt cactcctcag attgactcgt	1200
ccagagtaag aagtcacatc tgtagccacc caggggtgtg caagacttac tttaaaagt	1260
cccactgaa gggccacgtg aggcacacaca caggggaaaa accttcagc tgcagctgga	1320
aaggctgtga aaggaggtt gctcgtctcg atgaactgtc cagacaccgg cggacacaca	1380
caggtgagaa gaagtttgc tgtcccatgt gtgaccgtcg gtttatgagg agcgaccatt	1440
taaccaagca tgcccagcgc cacctatcag ccaagaagct gccaaactgg caaatggaag	1500
ttagcaagtt aaatgacatt gctctgcctc cgaccctcgc ttccgcacag tgacggccag	1560
aagatggaga cgcagaataa actttgttga gagtcaggag ccagtgatgg tgtcaagtgc	1620
ttctgcaagg ctgtggcctc ccaaaagggc ctaaagtaga agccctggcc tgggggaggc	1680
cccgcctggg tgaatgaca agaagtgctt cagccacagg caggtcacag aggacagggc	1740
tcagttctta ccacagagag agaggagaac ctttttattc ctcccttatt ttagtctgga	1800
aagtttcggc tgaggtgagc gcagcacagg ttttgaatca catacatt ggggactttg	1860
tttttgccat ttatacttga gaccagcttt gcagtgtgat ttttcaaag gatttggttc	1920
aagaatatag aggctggaaa ttacgggtaca gaaatggagc tagaaaatga gtttgtgta	1980
cacagagatg tcatctctc cttaggttat cttgtttctt attcctagtc tttccagtca	2040
aatccgtgga tgtagctaag tatactctaaa actcattttt ccactattgt tggattttga	2100
agttgaacag ctgtacattg ttgtggggga gccaaaggat tggaaacctc attaattdaa	2160
ttgcttgtaa atgcagctaa aattctctt tggcattttg ttttgaagt ttaggcattt	2220
tactctactt tagattttag tttgcttgca gttttttgtg tagattttaa aattgtatac	2280
caatgtgtt tctgtaggct taaaatacac tgcactttgt ttagaaaaa atctggagat	2340
gaaaaatgt attataaaga agagatgtca agaatttgag ataactcctt gagaaagtg	2400
gctttatgtc atcagcaaag gacacttaac gtcaagcata cactgtggtt tttttgtttt	2460
tttgtttttt tttttcaaat tagaaagttt aatgaccgtt acagatggac agtgtctttt	2520
tatttatagg agttttcag gatgtcacag tagataggta ggaaaattgt tattagaaca	2580
ttcgtctota ccttgaaaag gatgttaatg tggatcatgt cttagacca cagtgtctgg	2640
gcactctgga aactccgaga cttttttaaa gtgtcatgat gtgatcacac ctgcagtttg	2700
gggcatcgaa tccagggcct tgcagtctt ctgtaagagc tctcatcgct gacctgtatc	2760
ccccgcaaga gcaatgactt ttgctaacag tattttcttt ctgttgtaaa gtggacagat	2820
gatacacttg gtcgcaagg taaattattc aaaatccaca gtgaaaacct caccacactt	2880
tcccatttaa actatttcca tatctcagag gtttctgaca tgcaacttg aaccctgaa	2940
agaagagttt tcttaaaaat tataaaaaat cacgagttac aattgcaca atatttttg	3000
ttgaacttta taccttgttt acaataaaga cttttctttg gtata	3045

-continued

```

<210> SEQ ID NO 10
<211> LENGTH: 479
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
<220> FEATURE:
<223> OTHER INFORMATION: mouse transforming growth factor-beta (TGFB)
        inducible early growth response (TIEG) homolog

<400> SEQUENCE: 10

Met Leu Asn Phe Gly Ala Ser Leu Gln Gln Ala Ser Glu Gly Lys Met
  1           5           10
Glu Leu Ile Ser Glu Lys Pro Arg Glu Gly Met His Pro Trp Asp Lys
  20           25           30
Ala Glu Gln Ser Asp Phe Glu Ala Val Glu Ala Leu Met Ser Met Ser
  35           40           45
Cys Asp Trp Lys Ser His Phe Lys Lys Tyr Leu Glu Asn Arg Pro Val
  50           55           60
Thr Pro Val Ser Asp Thr Ser Glu Asp Asp Ser Leu Leu Pro Gly Thr
  65           70           75           80
Pro Asp Leu Gln Thr Val Pro Ala Phe Cys Leu Thr Pro Pro Tyr Ser
  85           90           95
Pro Ser Asp Phe Glu Pro Ser Gln Gly Ser Asn Leu Thr Ala Ser Ala
  100          105          110
Pro Ser Thr Gly His Phe Lys Ser Phe Ser Asp Ala Ala Lys Pro Pro
  115          120          125
Gly Ala Thr Pro Phe Lys Glu Glu Lys Asn Pro Leu Ala Ala Pro
  130          135          140
Pro Leu Pro Lys Ala Gln Ala Thr Ser Val Ile Arg His Thr Ala Asp
  145          150          155          160
Ala Gln Leu Cys Asn His Gln Ser Cys Pro Val Lys Ala Ala Ser Ile
  165          170          175
Leu Asn Tyr Gln Asp Asn Ser Phe Arg Arg Arg Thr His Gly Asn Val
  180          185          190
Glu Ala Thr Arg Lys Asn Ile Pro Cys Ala Ala Val Ser Pro Asn Arg
  195          200          205
Ser Lys Pro Glu Pro Ser Thr Val Ser Asp Gly Asp Glu Lys Ala Gly
  210          215          220
Ala Ala Leu Tyr Asp Phe Ala Val Pro Ser Ser Glu Thr Val Ile Cys
  225          230          235          240
Arg Ser Gln Pro Ala Pro Ser Ser Pro Val Gln Lys Ser Val Leu Val
  245          250          255
Ser Ser Pro Thr Val Ser Thr Gly Gly Val Pro Pro Leu Pro Val Ile
  260          265          270
Cys Gln Met Val Pro Leu Pro Ala Asn Asn Ser Leu Val Ser Thr Val
  275          280          285
Val Pro Ser Thr Pro Pro Ser Gln Pro Pro Ala Val Cys Ser Pro Val
  290          295          300
Leu Phe Met Gly Thr Gln Val Pro Glu Gly Thr Val Val Phe Val Val
  305          310          315          320
Pro Gln Pro Val Val Gln Ser Pro Arg Pro Pro Val Val Ser Pro Ser
  325          330          335
Gly Thr Arg Leu Ser Pro Ile Ala Pro Ala Pro Gly Phe Ser Pro Ser

```

-continued

340	345	350
Ala Ala Arg Val Thr Pro Gln Ile Asp Ser Ser Arg Val Arg Ser His		
355	360	365
Ile Cys Ser His Pro Gly Cys Gly Lys Thr Tyr Phe Lys Ser Ser His		
370	375	380
Leu Lys Ala His Val Arg Thr His Thr Gly Glu Lys Pro Phe Ser Cys		
385	390	395
Ser Trp Lys Gly Cys Glu Arg Arg Phe Ala Arg Ser Asp Glu Leu Ser		
405	410	415
Arg His Arg Arg Thr His Thr Gly Glu Lys Lys Phe Ala Cys Pro Met		
420	425	430
Cys Asp Arg Arg Phe Met Arg Ser Asp His Leu Thr Lys His Ala Arg		
435	440	445
Arg His Leu Ser Ala Lys Lys Leu Pro Asn Trp Gln Met Glu Val Ser		
450	455	460
Lys Leu Asn Asp Ile Ala Leu Pro Pro Thr Pro Ala Ser Ala Gln		
465	470	475

<210> SEQ ID NO 11
 <211> LENGTH: 3115
 <212> TYPE: DNA
 <213> ORGANISM: Rattus norvegicus
 <220> FEATURE:
 <223> OTHER INFORMATION: rat transforming growth factor-beta (TGFB)
 inducible early growth response (TIEG) homolog
 cDNA
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (316)..(1758)
 <223> OTHER INFORMATION: TIEG homolog

<400> SEQUENCE: 11

```

ggagggaaca cggtgacctg ggttgtgtac acgctccact gacagagctt cttgcagccg      60
ggcagccgct gatcacgcgt gccccgccca gccattggc tgaggcctca cacacctttg      120
ccgctgattg gtcgatacca cgccccgctt tccacccgcg ccgcgccgcg gggcaggcag      180
agcggctacc tgcattggga gggggggcag acagcgcgga gcgcggtggc gtcgacgtct      240
agtgtctcag tgctccgtct gtggctaggg agccagcagc caagcagctc gcgacctacg      300
gccaggcagc caacctatgt taacttcggc gcttctctcc agcaagcttc ggaggggaaa      360
atggaactaa tttctgaaaa gtccaaagag ggagcacatc cctgggacaa agctgaacag      420
agtgactttg aagctgtgga agcgtcatg tccatgagct gcgactggaa gtctcacttc      480
aagaaatacc ttgaaaacag gcctgtcaca ccagtgtctg atacctcaga ggaagacagt      540
ttgcttcacg ggacgcctga ccttcagacc gtcccagcat tctgtttgac tccaccttac      600
agccccctctg acttcgaacc ctgcgagggg tcaaactctga cggcaccagc gccacctact      660
ggtcacttca gatcgtctct cgatgctgcc aagcctccca gcatcgctcc ttttaaagag      720
gaggaaaaga gccattagc tgccccctcc ctccctaagg ctcaagccac cagtgtcctc      780
cgtcacacag ctgatgccca actctgtaac caccagtcct gccctgtgaa agcagctagc      840
atcctcaact accaggacaa ttctttccgg agaagaacct acataaatgt cgaggctact      900
cgaaagaaca taccctgtgc tgctgtgtca ccaaacagac ccaagcctga gccagcaca      960
cggcggaatg gcgctgagaa ggccggcact gcgccgtatg actttgtgtg gccttctca     1020
    
```

-continued

gagacagtca tctgtaggtc ttctcagcca gctcccacgt ccccagtaca gaagtcagta	1080
ctgatgtott cacctacagt gtccactggg ggagtgccac ccttgctgt catctgccag	1140
atggttcccc ttccagcaaa caactctctc gttaccacag ttgtcccag cagtccgccc	1200
agccagccac ccgctgtctg ctcacccgtg ttgttcatgg gcaactcaggt gcccaggggc	1260
acggctcatgt ttgtgttacc ccagcccgtt gtacagagcc caaagcctcc agtggtgagc	1320
cccaatggca ccagactgtc tcccattgcc cctgcccctg gattctctcc ttcagcagca	1380
agggttactc cccagattga ttcgtccaga gtaagaagtc acatctgtag ccaccagga	1440
tgtggcaaga cttactttaa aagttcccat ctgaaggccc acgtgagaac acacacaggg	1500
gaaaaacott tcagctgcag ctggaaaggc tgtgaaagga ggtttgcctg ttccgatgaa	1560
ctgtccagac accggcggac acacacaggt gagaagaaat ttgcctgtcc catgtgtgac	1620
cgtcggttca tgagaagtga ccatttaacc aaacatgccc ggcgacacct atcagccaag	1680
aagctgcaa actggcaaat ggaagtgagc aagttaaatg acatcgctct gcctccggcc	1740
acagcgtcgg cacagtgcg gccagaagat ggagactcag aataaacttt ggtcagagtc	1800
gagccggtga ccgtgtcaaa tgcttctgca agtctgccc cctccaacag ggctaaagc	1860
agaagcccct ggctggggg gaggcccagc ctgggttaga tgacgagaag tgcttcagcc	1920
acaggcagtt cacagaaggg cggggttcat ttcttaccac gtagagagat gagaacactc	1980
gtgtcccttg agtagttttg gaaggtttcg gatgagatga gcacagcaca ggttttgaat	2040
cacatacaca tcaggtactt tgtttttggc atttatactt gagaccagct ttgcagtgtg	2100
attcttttaa agcgttggtt togagaatat agaggctgaa attacggtac agaaatgcag	2160
ctagaaagtg agttttgttc atcttctctc agagttatct tgcttattat tccatgtctt	2220
tacagtcaac tccgtggcag tggctaagta tagaaaacta tcatttcttc cactgttggt	2280
ggtatttgag gttgaacagc tgtacattgc taaggggagc ccaaggattg gaactctcat	2340
taatttaatt gcttggaagt gcagctaaag ttcttttttt tggccttttg ttttgaaagt	2400
ttaggcattt tactctacct tagatattag ttgcttgca gttttttgtg tagatttgaa	2460
aactgtacac caatgtgttt cctatagact taagatagc tgccgtttgt ttagaaaaca	2520
atctggcagt gaaaatatgt gtcgtaaaga agagatgtca agaatttgag atgactcctt	2580
gagaaagatg gcttatgtca tcagcaaaag acacttacgt cagacgtaca ctgtttctat	2640
caaatcagga agttagtgc catcacagat ggacaggtgt ccttttattt ataggagttt	2700
gtcaggatgt ctgagtagat aggtaggaaa ttgttactag aacattcact tccgacttga	2760
caaggatatt catggtgcag atacgttctt cctgcccctg cagtgtctgg gtgtccggga	2820
aactccgaga ctttaaagtg tcataatgtg gtcacacctg cagttctggg gatccagagc	2880
cttgcttggt ttctgtaagc tctgtcaccg agttgtatcc ccacctgaaa gcagtgactt	2940
ttgctaacag tatttctgtt ctggtgtaaa gtggatagaa gacacacttg gtagtaaaag	3000
taaattattc aaagtccaca gtgaaaaact caccatacct tcccatttaa accgtttctg	3060
tatctcatag gtttctgaca tgcaaaactg aacccttgaa aaaaagagtt ttctt	3115

<210> SEQ ID NO 12

<211> LENGTH: 480

<212> TYPE: PRT

<213> ORGANISM: Rattus norvegicus

<220> FEATURE:

-continued

<223> OTHER INFORMATION: rat transforming growth factor-beta (TGFB)
inducible early growth response (TIEG) homolog

<400> SEQUENCE: 12

```

Met Leu Asn Phe Gly Ala Ser Leu Gln Gln Ala Ser Glu Gly Lys Met
 1           5           10           15
Glu Leu Ile Ser Glu Lys Ser Lys Glu Gly Ala His Pro Trp Asp Lys
 20           25           30
Ala Glu Gln Ser Asp Phe Glu Ala Val Glu Ala Leu Met Ser Met Ser
 35           40           45
Cys Asp Trp Lys Ser His Phe Lys Lys Tyr Leu Glu Asn Arg Pro Val
 50           55           60
Thr Pro Val Ser Asp Thr Ser Glu Glu Asp Ser Leu Leu Pro Gly Thr
 65           70           75
Pro Asp Leu Gln Thr Val Pro Ala Phe Cys Leu Thr Pro Pro Tyr Ser
 85           90           95
Pro Ser Asp Phe Glu Pro Ser Gln Gly Ser Asn Leu Thr Ala Pro Ala
100           105           110
Pro Pro Thr Gly His Phe Arg Ser Leu Ser Asp Ala Ala Lys Pro Pro
115           120           125
Ser Ile Ala Pro Phe Lys Glu Glu Glu Lys Ser Pro Leu Ala Ala Pro
130           135           140
Pro Leu Pro Lys Ala Gln Ala Thr Ser Val Ile Arg His Thr Ala Asp
145           150           155           160
Ala Gln Leu Cys Asn His Gln Ser Cys Pro Val Lys Ala Ala Ser Ile
165           170           175
Leu Asn Tyr Gln Asp Asn Ser Phe Arg Arg Arg Thr His Ile Asn Val
180           185           190
Glu Ala Thr Arg Lys Asn Ile Pro Cys Ala Ala Val Ser Pro Asn Arg
195           200           205
Pro Lys Pro Glu Pro Ser Thr Ala Ala Asn Gly Ala Glu Lys Ala Gly
210           215           220
Thr Ala Pro Tyr Asp Phe Ala Val Pro Ser Ser Glu Thr Val Ile Cys
225           230           235           240
Arg Ser Ser Gln Pro Ala Pro Thr Ser Pro Val Gln Lys Ser Val Leu
245           250           255
Met Ser Ser Pro Thr Val Ser Thr Gly Gly Val Pro Pro Leu Pro Val
260           265           270
Ile Cys Gln Met Val Pro Leu Pro Ala Asn Asn Ser Leu Val Thr Thr
275           280           285
Val Val Pro Ser Ser Pro Pro Ser Gln Pro Pro Ala Val Cys Ser Pro
290           295           300
Val Leu Phe Met Gly Thr Gln Val Pro Lys Gly Thr Val Met Phe Val
305           310           315           320
Val Pro Gln Pro Val Val Gln Ser Pro Lys Pro Pro Val Val Ser Pro
325           330           335
Asn Gly Thr Arg Leu Ser Pro Ile Ala Pro Ala Pro Gly Phe Ser Pro
340           345           350
Ser Ala Ala Arg Val Thr Pro Gln Ile Asp Ser Ser Arg Val Arg Ser
355           360           365
His Ile Cys Ser His Pro Gly Cys Gly Lys Thr Tyr Phe Lys Ser Ser
370           375           380

```


-continued

```

cgcacacagg agaaaagcct ttcagctgta gctggaaggg ttgtgaaagg aggtttgccc 1320
gttctgatga actgtccaga cacaggcgaa cccacacggg tgagaagaaa tttgcgtgcc 1380
ccatgtgtga ccggcggttc atgaggagtg accatttgac caagcatgcc cggcgccatc 1440
tatacgccaa gaagctacca aactggcaga tggaagtgag caagctaaat gacattgctc 1500
tacctccaac ccctgctccc acacagtgac agaccggaag gtgaagagtc agaactaact 1560
ttggtctcag cgggagccag tgggtgatgta aaaatgcttc cactgcaagt ctgtggcccc 1620
acaacgtggg cttaaagcag aagccccaca gcctggcacg aaggccccgc ctgggttagg 1680
tgactaaaag ggcttcggcc acaggcaggt cacagaaagg caggtttcat ttcttatcac 1740
ataagagaga tgagaaagct tttattcctt tgaatatttt ttgaaggttt cagatgaggt 1800
caacacaggt agcacagatt ttgaatctgt gtgcatattt gttactttac ttttctgttt 1860
tatacttgag accaactttt caatgtgatt cttctaaagc actggtttca agaatatgga 1920
agctggaagg aaataacat tacggtacag acatggagat gtaaatgag tttgtattat 1980
tacaaatatt gtcatctttt tctagagtta tcttctttat tattcctagt ctttccagtc 2040
aacatcgtgg atgtagtgat taaatatatc tagaactatc atttttacac tattgtgaat 2100
atltggaatt gaacgactgt atattgctaa gagggcccaa agaattgaa tcctccttaa 2160
tttaattgct ttgaagcata gctacaattt gtttttgcat ttttgttttg aaagttaac 2220
aaatgactgt atctaggcat ttcattatgc tttgaacttt agtttgctcg cagtttcttg 2280
tgtagatttg aaaattgtat accaatgtgt tttctgtaga ctctaagata cactgcaact 2340
tgtttagaaa aaaaactgaa gatgaaatat atattgtaaa gaagggatat taagaatctt 2400
agataacttc ttgaaaaaga tggcttatgt catcagtaaa gtacctttat gttatgagga 2460
tataatgtgt gctttattga attagaaaaat tagtgaccat tattcacagg tggacaaaatg 2520
ttcgtcctgt taatttatag gagttttttg gggatgtgga ggtagttggg tagaaaaatt 2580
attagaacat tcacttttgt taacagtatt tctcttttat tctgttatat agtggatgat 2640
atacacagtg gcaaaaacaaa agtcattatgc ttaaaatata tagtgaaaaa tgtcactata 2700
tcttccatt taacattggt tttgtatatt ggggttagat ttctgacatc aaaacttgga 2760
cccttggaag acaaaaagttt taattaaaaa aaatccttgt gacttacaat ttgcacaata 2820
tttcttttgt tgtactttat atcttgttta caataaagaa ttccttttgg ca 2872

```

```

<210> SEQ ID NO 14
<211> LENGTH: 480
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> OTHER INFORMATION: human transforming growth factor-beta (TGFB)
inducible early growth response (TIEG) splice
variant

```

```

<400> SEQUENCE: 14

```

```

Met Leu Asn Phe Gly Ala Ser Leu Gln Gln Thr Ala Glu Glu Arg Met
 1           5           10          15
Glu Met Ile Ser Glu Arg Pro Lys Glu Ser Met Tyr Ser Trp Asn Lys
          20          25          30
Thr Ala Glu Lys Ser Asp Phe Glu Ala Val Glu Ala Leu Met Ser Met
 35          40          45

```

-continued

Ser Cys Ser Trp Lys Ser Asp Phe Lys Lys Tyr Val Glu Asn Arg Pro
 50 55 60

Val Thr Pro Val Ser Asp Leu Ser Glu Glu Glu Asn Leu Leu Pro Gly
 65 70 75 80

Thr Pro Asp Phe His Thr Ile Pro Ala Phe Cys Leu Thr Pro Pro Tyr
 85 90 95

Ser Pro Ser Asp Phe Glu Pro Ser Gln Val Ser Asn Leu Met Ala Pro
 100 105 110

Ala Pro Ser Thr Val His Phe Lys Ser Leu Ser Asp Thr Ala Lys Pro
 115 120 125

His Ile Ala Ala Pro Phe Lys Glu Glu Glu Lys Ser Pro Val Ser Ala
 130 135 140

Pro Lys Leu Pro Lys Ala Gln Ala Thr Ser Val Ile Arg His Thr Ala
 145 150 155 160

Asp Ala Gln Leu Cys Asn His Gln Thr Cys Pro Met Lys Ala Ala Ser
 165 170 175

Ile Leu Asn Tyr Gln Asn Asn Ser Phe Arg Arg Arg Thr His Leu Asn
 180 185 190

Val Glu Ala Ala Arg Lys Asn Ile Pro Cys Ala Ala Val Ser Pro Asn
 195 200 205

Arg Ser Lys Cys Glu Arg Asn Thr Val Ala Asp Val Asp Glu Lys Ala
 210 215 220

Ser Ala Ala Leu Tyr Asp Phe Ser Val Pro Ser Ser Glu Thr Val Ile
 225 230 235 240

Cys Arg Ser Gln Pro Ala Pro Val Ser Pro Gln Gln Lys Ser Val Leu
 245 250 255

Val Ser Pro Pro Ala Val Ser Ala Gly Gly Val Pro Pro Met Pro Val
 260 265 270

Ile Cys Gln Met Val Pro Leu Pro Ala Asn Asn Pro Val Val Thr Thr
 275 280 285

Val Val Pro Ser Thr Pro Pro Ser Gln Pro Pro Ala Val Cys Pro Pro
 290 295 300

Val Val Phe Met Gly Thr Gln Val Pro Lys Gly Ala Val Met Phe Val
 305 310 315 320

Val Pro Gln Pro Val Val Gln Ser Ser Lys Pro Pro Val Val Ser Pro
 325 330 335

Asn Gly Thr Arg Leu Ser Pro Ile Ala Pro Ala Pro Gly Phe Ser Pro
 340 345 350

Ser Ala Ala Lys Val Thr Pro Gln Ile Asp Ser Ser Arg Ile Arg Ser
 355 360 365

His Ile Cys Ser His Pro Gly Cys Gly Lys Thr Tyr Phe Lys Ser Ser
 370 375 380

His Leu Lys Ala His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Ser
 385 390 395 400

Cys Ser Trp Lys Gly Cys Glu Arg Arg Phe Ala Arg Ser Asp Glu Leu
 405 410 415

Ser Arg His Arg Arg Thr His Thr Gly Glu Lys Lys Phe Ala Cys Pro
 420 425 430

Met Cys Asp Arg Arg Phe Met Arg Ser Asp His Leu Thr Lys His Ala
 435 440 445

Arg Arg His Leu Ser Ala Lys Lys Leu Pro Asn Trp Gln Met Glu Val

-continued

450	455	460
Ser Lys Leu Asn Asp Ile Ala Leu Pro Pro Thr Pro Ala Pro Thr Gln		
465	470	475
		480

<210> SEQ ID NO 15
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence:
 hexahistidine (His) affinity tag

<400> SEQUENCE: 15

His His His His His His
 1 5

<210> SEQ ID NO 16
 <211> LENGTH: 200
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence:poly-Gly
 flexible linker
 <220> FEATURE:
 <221> NAME/KEY: MOD_RES
 <222> LOCATION: (6)..(200)
 <223> OTHER INFORMATION: Gly residues from position 6 to 200 may be
 present or absent

<400> SEQUENCE: 16

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 1 5 10 15
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 20 25 30
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 35 40 45
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 50 55 60
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 65 70 75 80
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 85 90 95
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 100 105 110
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 115 120 125
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 130 135 140
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 145 150 155 160
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 165 170 175
 Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
 180 185 190
 Gly Gly Gly Gly Gly Gly Gly Gly Gly
 195 200

1. A method for identifying an agent for treating a diabetic or pre-diabetic individual, the method comprising the steps of:

(i) contacting an agent to a mixture comprising a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a nucleic acid encoding SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6; and

(ii) selecting an agent that modulates the expression or activity of the polypeptide or that binds to the polypeptide, thereby identifying an agent for treating a diabetic or pre-diabetic individual.

2. The method of claim 1, the method further comprising selecting an agent that modulates insulin sensitivity.

3. The method of claim 1, wherein step (ii) comprises selecting an agent that modulates expression of the polypeptide.

4. The method of claim 1, wherein step (ii) comprises selecting an agent that modulates the activity of the polypeptide.

5. The method of claim 1, wherein step (ii) comprises selecting an agent that specifically binds to the polypeptide.

6. The method of claim 1, wherein the polypeptide is expressed in a cell and the cell is contacted with the agent.

7.-16. (canceled)

17. A method of diagnosing an individual who has Type 2 diabetes or is prediabetic, the method comprising,

detecting in a sample from the individual the level of a polypeptide or the level of a polynucleotide encoding the polypeptide, wherein the polynucleotide hybridizes under stringent conditions to a nucleic acid encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6,

wherein a modulated level of the polypeptide or polynucleotide in the sample compared to a level of the polypeptide or polynucleotide in either a lean individual or a previous sample from the individual indicates that the individual is diabetic or prediabetic.

18. The method of claim 17, wherein the detecting step comprises contacting the sample with an antibody that specifically binds to the polypeptide.

19. The method of claim 17, wherein the detecting step comprises quantifying mRNA encoding the polypeptide.

20. The method of claim 19, wherein the mRNA is reverse transcribed and amplified in a polymerase chain reaction.

21. The method of claim 17, wherein the sample is a blood, urine or tissue sample.

22.-26. (canceled)

* * * * *

专利名称(译)	诊断和治疗糖尿病和胰岛素抵抗的方法		
公开(公告)号	US20060074018A1	公开(公告)日	2006-04-06
申请号	US10/516635	申请日	2003-06-04
[标]申请(专利权)人(译)	麦它波莱克斯股份有限公司		
申请(专利权)人(译)	METABOLEX INC		
当前申请(专利权)人(译)	METABOLEX INC.		
[标]发明人	BERNARD ALLAN GREGOIRE FRANCINE LAVAN BRIAN MOODIE SHONNA		
发明人	BERNARD, ALLAN GREGOIRE, FRANCINE LAVAN, BRIAN MOODIE, SHONNA		
IPC分类号	A61K38/22 G01N33/53 C07K14/47 C07K14/475 G01N33/74		
CPC分类号	A61K2039/505 C07K14/47 C07K14/475 G01N33/6893 G01N33/74 G01N2500/00 G01N2500/04 G01N2800/042		
优先权	60/386331 2002-06-05 US 60/386085 2002-06-04 US		
外部链接	Espacenet USPTO		

摘要(译)

本发明提供了用于诊断和治疗糖尿病和胰岛素抗性的组合物和方法。特别地，本发明提供了鉴定本发明多核苷酸或多肽的调节剂并使用那些调节剂治疗糖尿病的方法，以及通过测量患者中本发明的多核苷酸或多肽的水平来诊断糖尿病的方法。

