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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 July 2001 (12.07.2001)

(41) Application Number
US00/00360

(42) International Application Number
PCT/US01/00360

(71) Applicant (for all designated States except US):
GENOPTERA, LLC [US/US]; PO. Box 511, 170 Harbor Way, South San Francisco, CA 94083-0511 (US).

(72) Inventors and

Published:
— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACIDS AND POLYPEPTIDES OF INVERTEBRATE ENZYMES AND METHODS OF USE

(57) Abstract: The invention provides isolated invertebrate enzyme-encoding nucleic acids, and proteins encoded thereby, including lysine-ketoglutarate reductase/saccharopine dehydrogenase (LKR/SDH) and peptidyl-alpha-hydroxylglycine alpha-amidating lyase (PAL). The subject nucleic acid and protein can be used to genetically modify metazoan invertebrate organisms, such as insects and worms, or cultured cells, resulting in expression or mis-expression of a subject protein. The genetically modified organisms or cells can be used in screening assays to identify candidate compounds which are potential pesticidal agents or therapeutics that interact with a subject protein. They can also be used in methods for studying activity of a subject protein and identifying other genes that modulate the function of, or interact with, a subject gene.
NUCLEIC ACIDS AND POLYPEPTIDES OF INVERTEBRATE ENZYMES
AND METHODS OF USE

BACKGROUND OF THE INVENTION

Peptide alpha-amidation is a widespread, often essential post-translational modification shared by many bioactive peptides. Bioactive peptide include hormones, neuropeptides and paracrine agents; generally functioning as signaling molecules across great distances in the body. Examples include substance P, neuropeptide Y, thyrotropin and gonadotropin releasing hormones, oxytocin and vasopressin, cholecystokinin and gastrin, calcitonin, many of the snail conotoxins, locust adipokinetic hormone, Aplysia egg-laying hormone, and pyro-Glu-Tyr-Pro-NH$_2$ from alfalfa. For nearly all alpha-amidated peptides, the alpha-amide group is required for full biological activity. This amidation of the C-terminal residue of peptides is generally accomplished by the products of a single gene encoding a multifunctional protein, peptidylglycine alpha-amidating monoxygenase (PAM) (for review, see Eipper et al., Protein Science (1993) 2:489-497).

PAM has two catalytic domains that work sequentially to produce the final amidated product. The enzyme catalyzing "reaction 1" is peptidylglycine alpha-hydroxylating monoxygenase (PHM). This first reaction is rate limiting and requires copper, molecular oxygen and ascorbate. The reaction intermediate produced is quite stable at pH ~5, which is what is found in secretory granules where most of this activity is localized. "Reaction 2" is catalyzed by Peptidyl-alpha- hydroxyglycine alpha-amidating lyase (PAL). This step proceeds spontaneously at alkaline pH. PAL activity is abolished by EDTA, indicating the presence of a catalytic divalent cation. This activity can be restored through the addition of several different divalent cations. There is no allosteria between the two enzymes in the bifunctional forms which have been studied. The two enzymes have also been expressed individually and shown to have no loss of specific activity. It has also been shown that a certain percentage of PAM is proteolytically processed in cells, separating the PHM and PAL enzymes. The PHM enzyme then becomes a soluble protein, and is stored in secretory granules, while the PAL domain remains membrane bound. In Drosophila, these enzymes have been found to be the products of different genes, as PHM and PAL are expressed as separate polypeptides (Kolhekar et al., J. Neuroscience (1997) 17:1363-1376).

Lysine-ketoglutarate reductase/saccharopine dehydrogenase (LKR/SDH) is a principal enzyme in the catabolism of lysine. Defects in the enzyme lead to hyperlysinemia. Familial hyperlysinemia is an autosomal recessive disorder in humans. Regulation of lysine, cysteine and methionine levels is very important to the health and productivity of crop plants, and thus this pathway is also of considerable interest in agriculture.
LKR/SDH is activated by Ca\textsuperscript{2+}, high salt concentrations, organic solvents and Mg\textsuperscript{2+}. Enzyme isolated from maize and treated with elastase separated into two distinct enzymes: a 65 kD LKR polypeptide, and a 57 kD SDH polypeptide. The LKR domain retained the Ca\textsuperscript{2+} activation property, but was no longer activated by high salt concentrations, implying that the native enzyme is normally inhibited and when activated undergoes a conformational alteration to expose the catalytic domain for substrate binding.

Pesticide development has traditionally focused on the chemical and physical properties of the pesticide itself, a relatively time-consuming and expensive process. As a consequence, efforts have been concentrated on the modification of pre-existing, well-validated compounds, rather than on the development of new pesticides. There is a need in the art for new pesticidal compounds that are safer, more selective, and more efficient than currently available pesticides. The present invention addresses this need by providing novel pesticide targets from invertebrates such as the fruit fly *Drosophila melanogaster*, and by providing methods of identifying compounds that bind to and modulate the activity of such targets.

**SUMMARY OF THE INVENTION**

It is an object of the invention to provide isolated insect nucleic acids and proteins that are targets for pesticides. The isolated insect nucleic acid molecules provided herein are useful for producing insect proteins encoded thereby. The insect proteins are useful in assays to identify compounds that modulate a biological activity of the proteins, which assays identify compounds that may have utility as pesticides. It is an object of the present invention to provide invertebrate homologs of genes encoding enzymes that can be used in genetic screening methods to characterize pathways that such genes may be involved in, as well as other interacting genetic pathways. It is also an object of the invention to provide methods for screening compounds that interact with a subject invertebrate enzyme.

Compounds that interact with a subject invertebrate enzyme may have utility as therapeutics or pesticides. In some embodiments, a subject invertebrate enzyme is a lysine-ketoglutarate reductase/saccharopine dehydrogenase (LKR/SDH). In other embodiments, a subject invertebrate enzyme is a peptidyl-alpha-hydroxyglycine alpha amidating lyase (PAL). In some embodiments, the subject enzymes are from *Drosophila melanogaster*. In particular embodiments, subject enzymes are provided that are designated dmLKR/SDH, dmPAL, and dmPAL2.

These and other objects are provided by the present invention which concerns the identification and characterization of novel pesticide targets in insects, e.g., *Drosophila melanogaster*. Isolated nucleic acid molecules are provided that comprise nucleic acid sequences encoding subject proteins as well as novel fragments and derivatives thereof. Methods of using the isolated nucleic acid molecules and fragments of the invention as biopesticides are described, such as use of RNA interference methods.
that block a biological activity of a subject protein. Vectors and host cells comprising the subject nucleic acid molecules are also described, as well as metazoan invertebrate organisms (e.g. insects, coelomates and pseudocoelomates) that are genetically modified to express or mis-express a subject protein.

An important utility of the subject nucleic acids and proteins is that they can be used in screening assays to identify candidate compounds which are potential pesticidal agents or therapeutics that interact with subject proteins. Such assays typically comprise contacting a subject protein or fragment with one or more candidate molecules, and detecting any interaction between the candidate compound and the subject protein. The assays may comprise adding the candidate molecules to cultures of cells genetically engineered to express subject proteins, or alternatively, administering the candidate compound to a metazoan invertebrate organism genetically engineered to express a subject protein.

The genetically engineered metazoan invertebrate animals of the invention can also be used in methods for studying a biological activity of a subject protein. These methods typically involve detecting the phenotype caused by the expression or mis-expression of the subject protein. The methods may additionally comprise observing a second animal that has the same genetic modification as the first animal and, additionally has a mutation in a gene of interest. Any difference between the phenotypes of the two animals identifies the gene of interest as capable of modifying the function of the gene encoding the subject protein.

**DETAILED DESCRIPTION OF THE INVENTION**

It is an object of the invention to provide isolated insect nucleic acids and proteins that are targets for pesticides. The isolated insect nucleic acid molecules provided herein are useful for producing insect proteins encoded thereby. The insect proteins are useful in assays to identify compounds that modulate a biological activity of the proteins, which assays identify compounds that may have utility as pesticides. It is an object of the present invention to provide invertebrate homologs of subject genes that can be used in genetic screening methods to characterize pathways that such genes may be involved in, as well as other interacting genetic pathways. It is also an object of the invention to provide methods for screening compounds that interact with a subject protein. Compounds that interact with a subject protein may have utility as therapeutics or pesticides. In general, the subject proteins are enzymes. In some embodiments, a subject protein is a lysine-ketoglutarate reductase/saccharopine dehydrogenase (LKR/SDH). In other embodiments, a subject protein is a peptidyl-alpha-hydroxyglycine alpha amidating lyase (PAL). In some embodiments, the subject enzymes are from *Drosophila melanogaster*. In particular embodiments, subject enzymes are provided that are designated dmlKR/SDH, dmPAL, and dmPAL2.
These and other objects are provided by the present invention, which concerns the identification and characterization of novel pesticide targets in insects, e.g., *Drosophila melanogaster*. Isolated nucleic acid molecules are provided that comprise nucleic acid sequences encoding subject proteins as well as novel fragments and derivatives thereof. Methods of using the isolated nucleic acid molecules and fragments of the invention as biopesticides are described, such as use of RNA interference methods that block a biological activity of a subject protein. Vectors and host cells comprising the subject nucleic acid molecules are also described, as well as metazoan invertebrate organisms (e.g. insects, coelomates and pseudocoelomates) that are genetically modified to express or mis-express a subject protein.

An important utility of the subject nucleic acids and proteins is that they can be used in screening assays to identify candidate compounds which are potential pesticidal agents or therapeutics that interact with subject proteins. Such assays typically comprise contacting a subject protein or fragment with one or more candidate molecules, and detecting any interaction between the candidate compound and the subject protein. The assays may comprise adding the candidate molecules to cultures of cells genetically engineered to express subject proteins, or alternatively, administering the candidate compound to a metazoan invertebrate organism genetically engineered to express a subject protein.

The genetically engineered metazoan invertebrate animals of the invention can also be used in methods for studying a biological activity of a subject protein. These methods typically involve detecting the phenotype caused by the expression or mis-expression of the subject protein. The methods may additionally comprise observing a second animal that has the same genetic modification as the first animal and, additionally has a mutation in a gene of interest. Any difference between the phenotypes of the two animals identifies the gene of interest as capable of modifying the function of the gene encoding the subject protein.

Systematic genetic analysis of subject proteins using invertebrate model organisms can lead to the identification and validation of pesticide targets directed to components of a pathway involving a subject protein. Model organisms or cultured cells that have been genetically engineered to express a subject protein can be used to screen candidate compounds for their ability to modulate expression or activity of a subject nucleic acid or protein, and thus are useful in the identification of new drug targets, therapeutic agents, diagnostics and prognostics useful in the treatment of disorders associated with abnormal enzyme levels and/or function, e.g., disorders associated with lysine catabolism. Additionally, these invertebrate model organisms can be used for the identification and screening of pesticide targets directed to components of a pathway involving a subject protein.

The details of the conditions used for the identification and/or isolation of novel nucleic acids and proteins of the invention are described in the Examples section below. Various non-limiting embodiments of the invention, applications and uses of these novel genes and proteins are discussed in the following sections. The entire contents of all references, including patent applications, cited herein
are incorporated by reference in their entireties for all purposes. Additionally, the citation of a reference in the preceding background section is not an admission of prior art against the claims appended hereto.

For the purposes of the present application, singular forms "a", "an", and "the" include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to "an invertebrate receptor" includes large numbers of receptors, reference to "an agent" includes large numbers of agents and mixtures thereof, reference to "the method" includes one or more methods or steps of the type described herein.

DEFINITIONS

As used herein the term "isolated" is meant to describe a polynucleotide, a polypeptide, an antibody, or a host cell that is in an environment different from that in which the polynucleotide, the polypeptide, the antibody, or the host cell naturally occurs. As used herein, the term "substantially purified" refers to a compound (e.g., either a polynucleotide or a polypeptide or an antibody) that is removed from its natural environment and is at least 60% free, preferably 75% free, and most preferably 90% free from other components with which it is naturally associated.

The terms "polypeptide" and "protein", used interchangeably herein, refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. The term includes fusion proteins, including, but not limited to, fusion proteins with a heterologous amino acid sequence, fusions with heterologous and homologous leader sequences, with or without N-terminal methionine residues; immunologically tagged proteins; and the like.

A "host cell", as used herein, denotes microorganisms or eukaryotic cells or cell lines cultured as unicellular entities which can be, or have been, used as recipients for recombinant vectors or other transfer polynucleotides, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

By "transformation" is meant a permanent or transient genetic change induced in a cell following incorporation of new DNA (i.e., DNA exogenous to the cell). Genetic change can be accomplished either by incorporation of the new DNA into the genome of the host cell, or by transient or stable maintenance of the new DNA as an episomal element. Where the cell is a eukaryotic cell, a permanent genetic change is generally achieved by introduction of the DNA into the genome of the cell.
ISOLATED NUCLEIC ACID MOLECULES OF THE INVENTION

The invention provides isolated insect nucleic acid molecules comprising nucleotide sequences of invertebrate enzymes, particularly nucleic acid sequences of insect enzymes, and more particularly nucleic acid sequences of *Drosophila* enzymes, and methods of using these nucleic acid molecules.

The present invention provides isolated nucleic acid molecules that comprise nucleotide sequences encoding insect proteins that are potential pesticide targets. The isolated nucleic acid molecules have a variety of uses, e.g., as hybridization probes, e.g., to identify nucleic acid molecules that share nucleotide sequence identity; in expression vectors to produce the polypeptides encoded by the nucleic acid molecules; and to modify a host cell or animal for use in assays described hereinbelow.

Thus, the term "isolated nucleic acid sequence", as used herein, includes the reverse complement, RNA equivalent, DNA or RNA single- or double-stranded sequences, and DNA/RNA hybrids of the sequence being described, unless otherwise indicated.

The terms "polynucleotide" and "nucleic acid molecule", used interchangeably herein, refer to a polymeric forms of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The backbone of the polynucleotide can comprise sugars and phosphate groups (as may typically be found in RNA or DNA), or modified or substituted sugar or phosphate groups. Alternatively, the backbone of the polynucleotide can comprise a polymer of synthetic subunits such as phosphoramidites and thus can be an oligodeoxynucleoside phosphoramidate or a mixed phosphoramidate-phosphodiester oligomer. Peyrottes et al. (1996) *Nucl. Acids Res.* 24:1841-1848; Chaturvedi et al. (1996) *Nucl. Acids Res.* 24:2318-2323. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs, uracil, other sugars, and linking groups such as fluororibose and thioate, and nucleotide branches. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications included in this definition are caps, substitution of one or more of the naturally occurring nucleotides with an analog, and introduction of means for attaching the polynucleotide to proteins, metal ions, labeling components, other polynucleotides, or a solid support.

For hybridization probes, it may be desirable to use nucleic acid analogs, in order to improve the stability and binding affinity. A number of modifications have been described that alter the chemistry of the phosphodiester backbone, sugars or heterocyclic bases.

Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O-5'-$S$-
phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH2-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire phosphodiester backbone with a peptide linkage.

Sugar modifications are also used to enhance stability and affinity. The \( \alpha \)-anomer of deoxyribose may be used, where the base is inverted with respect to the natural \( \beta \)-anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without compromising affinity. Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxycytidine; 5-methyl-2'-deoxyctydine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxycytidine and deoxycytidine, respectively.

In addition to the fragments and derivatives of SEQ ID NOS:1, 3, and 5, as described in detail below, the invention includes the reverse complements thereof. Also, the subject nucleic acid sequences, derivatives and fragments thereof may be RNA molecules comprising the nucleotide sequence of SEQ ID NOS:1, 3, and 5 (or derivative or fragment thereof) wherein the base U (uracil) is substituted for the base T (thymine). The DNA and RNA sequences of the invention can be single- or double-stranded. Thus, the term “isolated nucleic acid sequence”, as used herein, includes the reverse complement, RNA equivalent, DNA or RNA single- or double-stranded sequences, and DNA/RNA hybrids of the sequence being described, unless otherwise indicated.

Fragments of the subject nucleic acid molecules can be used for a variety of purposes. Interfering RNA (RNAi) fragments, particularly double-stranded (ds) RNAi, can be used to generate loss-of-function phenotypes, or to formulate biopesticides (discussed further below). The subject nucleic acid fragments are also useful as nucleic acid hybridization probes and replication/amplification primers. Certain “antisense” fragments, i.e. that are reverse complements of portions of the coding sequence of any one of SEQ ID NOS:1, 3, and 5 have utility in inhibiting the function of a subject protein. The fragments are of length sufficient to specifically hybridize with the corresponding any one of SEQ ID NOS:1, 3, and 5. The fragments consist of or comprise at least 12, preferably at least 24, more preferably at least 36, and more preferably at least 96 contiguous nucleotides of any one of SEQ ID NOS:1, 3, and 5. When the fragments are flanked by other nucleic acid sequences, the total length of the combined nucleic acid sequence is less than 15 kb, preferably less than 10 kb or less than 5 kb, and more preferably less than 2 kb.

The subject nucleic acid sequences may consist solely of any one of SEQ ID NOS:1, 3, and 5 or fragments thereof. Alternatively, the subject nucleic acid sequences and fragments thereof may be joined to other components such as labels, peptides, agents that facilitate transport across cell membranes, hybridization-triggered cleavage agents or intercalating agents. The subject nucleic acid sequences and
fragments thereof may also be joined to other nucleic acid sequences (i.e. they may comprise part of larger sequences) and are of synthetic/non-natural sequences and/or are isolated and/or are purified, i.e. unaccompanied by at least some of the material with which it is associated in its natural state. Preferably, the isolated nucleic acids constitute at least about 0.5%, and more preferably at least about 5% by weight of the total nucleic acid present in a given fraction, and are preferably recombinant, meaning that they comprise a non-natural sequence or a natural sequence joined to nucleotide(s) other than that which it is joined to on a natural chromosome.

Derivative nucleic acid molecules of the subject nucleic acid molecules include sequences that hybridize to the nucleic acid sequence of any one of SEQ ID NOS:1, 3, and 5 under stringency conditions such that the hybridizing derivative nucleic acid is related to the subject nucleic acid by a certain degree of sequence identity. A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule. Stringency of hybridization refers to conditions under which nucleic acids are hybridizable. The degree of stringency can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. As used herein, the term "stringent hybridization conditions" are those normally used by one of skill in the art to establish at least a 90% sequence identity between complementary pieces of DNA or DNA and RNA. "Moderately stringent hybridization conditions" are used to find derivatives having at least 70% sequence identity. Finally, "low-stringency hybridization conditions" are used to isolate derivative nucleic acid molecules that share at least about 50% sequence identity with the subject nucleic acid sequence.

The ultimate hybridization stringency reflects both the actual hybridization conditions as well as the washing conditions following the hybridization, and it is well known in the art how to vary the conditions to obtain the desired result. Conditions routinely used are set out in readily available procedure texts (e.g., Current Protocol in Molecular Biology, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook et al., Molecular Cloning, Cold Spring Harbor (1989)). A preferred derivative nucleic acid is capable of hybridizing to any one of SEQ ID NOS:1, 3, and 5 under stringent hybridization conditions that comprise: prehybridization of filters containing nucleic acid for 8 hours overnight at 65°C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X Denhardt's solution, 0.05% sodium pyrophosphate and 100 µg/ml herring sperm DNA; hybridization for 18-20 hours at 65°C in a solution containing 6X SSC, 1X Denhardt's solution, 100 µg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65°C for 1 h in a solution containing 0.2X SSC and 0.1% SDS (sodium dodecyl sulfate).

Derivative nucleic acid sequences that have at least about 70% sequence identity with any one of SEQ ID NOS:1, 3, and 5 are capable of hybridizing to any one of SEQ ID NOS:1, 3, and 5 under
moderately stringent conditions that comprise: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μg/ml denatured salmon sperm DNA; hybridization for 18-20 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 μg/ml salmon sperm DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS.

Other preferred derivative nucleic acid sequences are capable of hybridizing to any one of SEQ ID NOS:1, 3, and 5 under low stringency conditions that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt’s solution, 10% dextran sulfate, and 20 μg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

As used herein, “percent (%) nucleic acid sequence identity” with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of nucleotides in the candidate derivative nucleic acid sequence identical with the nucleotides in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul et al., J. Mol. Biol. (1997) 215:403-410; http://blast.wustl.edu/blast/README.html; hereinafter referred to generally as “BLAST”) with all the search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A percent (%) nucleic acid sequence identity value is determined by the number of matching identical nucleotides divided by the sequence length for which the percent identity is being reported.

Derivatives of a subject nucleic acid molecule usually have at least 70% sequence identity, preferably at least 80% sequence identity, more preferably at least 85% sequence identity, still more preferably at least 90% sequence identity, and most preferably at least 95% sequence identity with any one of SEQ ID NOS:1, 3, and 5, or domain-encoding regions thereof.

In one preferred embodiment, the derivative nucleic acid encodes a polypeptide comprising an amino acid sequence set forth in any one of SEQ ID NOS:2, 4, and 6, or a fragment or derivative thereof as described further below. A derivative of a subject nucleic acid molecule, or fragment thereof, may comprise 100% sequence identity with any one of SEQ ID NOS:1, 3, and 5, but be a derivative thereof in the sense that it has one or more modifications at the base or sugar moiety, or phosphate backbone. Examples of modifications are well known in the art (Bailey, Ullmann’s Encyclopedia of Industrial
Chemistry (1998), 6th ed. Wiley and Sons). Such derivatives may be used to provide modified stability or any other desired property.

Another type of derivative of the subject nucleic acid sequences includes corresponding humanized sequences. A humanized nucleic acid sequence is one in which one or more codons has been substituted with a codon that is more commonly used in human genes. Preferably, a sufficient number of codons have been substituted such that a higher level expression is achieved in mammalian cells than what would otherwise be achieved without the substitutions. Tables are available in the art that show, for each amino acid, the calculated codon frequency in human genes for 1000 codons (Wada et al., Nucleic Acids Research (1990) 18(Suppl.):2367-2411). Similarly, other nucleic acid derivatives can be generated with codon usage optimized for expression in other organisms, such as yeasts, bacteria, and plants, where it is desired to engineer the expression of receptor proteins by using specific codons chosen according to the preferred codons used in highly expressed genes in each organism. Thus, a subject nucleic acid molecule in which the glutamic acid codon, GAA has been replaced with the codon GAG, which is more commonly used in human genes, is an example of a humanized nucleic acid molecule. A detailed discussion of the humanization of nucleic acid sequences is provided in U.S. Pat. No. 5,874,304 to Zolotukhin et al.

Specific nucleic acid molecules of the invention are discussed in detail below.

**dmLKR/SDH Nucleic Acid Molecules**

In some embodiments, the invention provides isolated nucleic acid molecules comprising nucleotide sequences encoding an invertebrate LKR/SDH. As described in the Examples below, a nucleic acid sequence (SEQ ID NO:1) was isolated from *Drosophila* that encodes a LKR/SDH homolog, hereinafter referred to as dmLKR/SDH.

In some embodiments, a dmLKR/SDH nucleic acid molecule comprises a nucleotide sequence of at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 300, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, at least about 1600, at least about 1700, at least about 1800, at least about 1900, at least about 2000, at least about 2100, at least about 2200, at least about 2300 at least about 2400, at least about 2500, at least about 2600, at least about 2700, at least about 2800, at least about 2900, at least about 3000, at least about 3100, or at least about 3200 contiguous nucleotides of the sequence set forth in SEQ ID NO:1, up to the entire sequence set forth in SEQ ID NO:1.

In other embodiments, a dmLKR/SDH nucleic acid molecule of the invention comprises a nucleotide sequence that encodes a polypeptide comprising at least about 6, at least about 10, at least
about 20, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, or at least about 900 contiguous amino acids of the sequence set forth in SEQ ID NO:2, up to the entire amino acid sequence as set forth in SEQ ID NO:2.

Derivative dmLKR/SDH nucleic acid sequences usually have at least 70% sequence identity, preferably at least 80% sequence identity, more preferably at least 85% sequence identity, still more preferably at least 90% sequence identity, and most preferably at least 95% sequence identity with SEQ ID NO:1, or domain-encoding regions thereof.

In one preferred embodiment, the derivative nucleic acid encodes a polypeptide comprising a dmLKR/SDH amino acid sequence of SEQ ID NO:2, or a fragment or derivative thereof as described further below under the subheading “dmLKR/SDH proteins”.

More specific embodiments of preferred dmLKR/SDH protein fragments and derivatives are discussed further below in connection with specific dmLKR/SDH proteins.

dmPAL Nucleic Acid Molecules

In some embodiments, the invention provides isolated nucleic acid molecules comprising nucleotide sequences encoding an invertebrate PAL. As described in the Examples below, a nucleic acid sequence (SEQ ID NO:3) was isolated from Drosophila that encodes a PAL homolog, hereinafter referred to as dmPAL.

In some embodiments, a dmPAL nucleic acid molecule comprises a nucleotide sequence of at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 300, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, at least about 1600, at least about 1700, at least about 1800, at least about 1900, at least about 2000, at least about 2100, at least about 2200, at least about 2300 at least about 2400, at least about 2500, or at least about 2600 contiguous nucleotides of the sequence set forth in SEQ ID NO:3, up to the entire sequence set forth in SEQ ID NO:3.

In other embodiments, a dmPAL nucleic acid molecule of the invention comprises a nucleotide sequence that encodes a polypeptide comprising at least about 6, at least about 10, at least about 20, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, or at least about 540 contiguous amino acids of the sequence set forth in SEQ ID NO:4, up to the entire amino acid sequence as set forth in SEQ ID NO:4.
Additional preferred fragments of SEQ ID NO:3 encode extracellular or intracellular domains which are located at approximately nucleotides 522-568, 621-1912, and 1965-2144.

Derivative dmPAL nucleic acid sequences usually have at least 70% sequence identity, preferably at least 80% sequence identity, more preferably at least 85% sequence identity, still more preferably at least 90% sequence identity, and most preferably at least 95% sequence identity with SEQ ID NO:3, or domain-encoding regions thereof.

In one preferred embodiment, the derivative nucleic acid encodes a polypeptide comprising a dmPAL amino acid sequence of SEQ ID NO:4, or a fragment or derivative thereof as described further below under the subheading "dmPAL proteins".

More specific embodiments of preferred dmPAL protein fragments and derivatives are discussed further below in connection with specific dmPAL proteins.

dmPAL2 Nucleic Acid Molecules

In some embodiments, the invention provides isolated nucleic acid molecules comprising nucleotide sequences encoding an invertebrate PAL2. As described in the Examples below, a nucleic acid sequence (SEQ ID NO:5) was isolated from Drosophila that encodes a PAL2 homolog, hereinafter referred to as dmPAL2.

In some embodiments, a dmPAL2 nucleic acid molecule comprises a nucleotide sequence of at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 300, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, at least about 1600, or at least about 1700, contiguous nucleotides of the sequence set forth in SEQ ID NO:5, up to the entire sequence set forth in SEQ ID NO:5.

In other embodiments, a dmPAL2 nucleic acid molecule of the invention comprises a nucleotide sequence that encodes a polypeptide comprising at least about 6, at least about 10, at least about 20, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, or at least about 400 contiguous amino acids of the sequence set forth in SEQ ID NO:6, up to the entire amino acid sequence as set forth in SEQ ID NO:6.

Additional preferred fragments of SEQ ID NO:5 encode NHL domains, which are located at approximately nucleotides 647-730, 800-886, 941-1027 and 1223-1321.

Derivative dmPAL2 nucleic acid sequences usually have at least 70% sequence identity, preferably at least 80% sequence identity, more preferably at least 85% sequence identity, still more preferably at least 90% sequence identity, and most preferably at least 95% sequence identity with SEQ ID NO:5, or domain-encoding regions thereof.
In one preferred embodiment, the derivative nucleic acid encodes a polypeptide comprising a dmPAL2 amino acid sequence of SEQ ID NO:6, or a fragment or derivative thereof as described further below under the subheading “dmPAL2 proteins”.

More specific embodiments of preferred dmPAL2 protein fragments and derivatives are discussed further below in connection with specific dmPAL2 proteins.

**Isolation, Production, and Expression of Subject Nucleic Acid Molecules**

The subject nucleic acid molecules, or fragments or derivatives thereof, may be obtained from an appropriate cDNA library prepared from any eukaryotic species that encodes a subject protein, such as vertebrates, preferably mammalian (e.g. primate, porcine, bovine, feline, equine, and canine species, etc.) and invertebrates, such as arthropods, particularly insects species (preferably Drosophila), acarids, crustacea, molluscs, nematodes, and other worms. An expression library can be constructed using known methods. For example, mRNA can be isolated to make cDNA which is ligated into a suitable expression vector for expression in a host cell into which it is introduced. Various screening assays can then be used to select for the gene or gene product (e.g. oligonucleotides of at least about 20 to 80 bases designed to identify the gene of interest, or labeled antibodies that specifically bind to the gene product). The gene and/or gene product can then be recovered from the host cell using known techniques.

Polymerase chain reaction (PCR) can also be used to isolate a subject nucleic acid molecule, where oligonucleotide primers representing fragmentary sequences of interest amplify RNA or DNA sequences from a source such as a genomic or cDNA library (as described by Sambrook *et al.*, *supra*). Additionally, degenerate primers for amplifying homologs from any species of interest may be used. Once a PCR product of appropriate size and sequence is obtained, it may be cloned and sequenced by standard techniques, and utilized as a probe to isolate a complete cDNA or genomic clone.

Fragmentary sequences of the subject nucleic acid molecules and derivatives thereof may be synthesized by known methods. For example, oligonucleotides may be synthesized using an automated DNA synthesizer available from commercial suppliers (e.g. Biosearch, Novato, CA; Perkin-Elmer Applied Biosystems, Foster City, CA). Antisense RNA sequences can be produced intracellularly by transcription from an exogenous sequence, e.g. from vectors that contain subject antisense nucleic acid sequences. Newly generated sequences may be identified and isolated using standard methods.

An isolated subject nucleic acid molecule can be inserted into any appropriate cloning vector, for example bacteriophages such as lambda derivatives, or plasmids such as pBR322, pUC plasmid derivatives and the Bluescript vector (Stratagene, San Diego, CA). Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, *etc.*, or into a transgenic animal such as a fly. The transformed cells can be cultured to generate large quantities of the subject nucleic acid. Suitable methods for isolating and producing the subject nucleic acid sequences are

The nucleotide sequence encoding a subject protein or fragment or derivative thereof, can be inserted into any appropriate expression vector for the transcription and translation of the inserted protein-coding sequence. Alternatively, the necessary transcriptional and translational signals can be supplied by the native subject gene and/or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence such as mammalian cell systems infected with virus (e.g. vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (e.g. baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. Expression of a subject protein may be controlled by a suitable promoter/enhancer element. In addition, a host cell strain may be selected which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired.

To detect expression of a subject gene product, the expression vector can comprise a promoter operably linked to a subject nucleic acid molecule, one or more origins of replication, and, one or more selectable markers (e.g. thymidine kinase activity, resistance to antibiotics, etc.). Alternatively, recombinant expression vectors can be identified by assaying for the expression of a subject gene product based on the physical or functional properties of a subject protein in in vitro assay systems (e.g. immunoassays).

A subject protein, fragment, or derivative may be optionally expressed as a fusion, or chimeric protein product (i.e. it is joined via a peptide bond to a heterologous protein sequence of a different protein). A chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other in the proper coding frame using standard methods and expressing the chimeric product. A chimeric product may also be made by protein synthetic techniques, e.g. by use of a peptide synthesizer.

Once a recombinant that expresses a subject gene sequence is identified, the gene product can be isolated and purified using standard methods (e.g. ion exchange, affinity, and gel exclusion chromatography; centrifugation; differential solubility; electrophoresis). The amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant and can thus be synthesized by standard chemical methods (Hunkapiller et al., Nature (1984) 310:105-111). Alternatively, native subject proteins can be purified from natural sources, by standard methods (e.g. immunoaffinity purification).
**ISOLATED POLYPEPTIDES OF THE INVENTION**

The invention further provides isolated polypeptides comprising or consisting of an amino acid sequence of any of SEQ ID NOS:2, 4, or 6, or fragments or derivatives thereof. Compositions comprising any of these proteins may consist essentially of a subject protein, fragments, or derivatives, or may comprise additional components (e.g. pharmaceutically acceptable carriers or excipients, culture media, carriers used in pesticide formulations, etc.).

A derivative of a subject protein typically shares a certain degree of sequence identity or sequence similarity with any one of SEQ ID NOS:2, 4, or 6, or a fragment thereof. As used herein, “percent (%) amino acid sequence identity” with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of amino acids in the candidate derivative amino acid sequence identical with the amino acid in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by BLAST (Altschul et al., supra) using the same parameters discussed above for derivative nucleic acid sequences. A % amino acid sequence identity value is determined by the number of matching identical amino acids divided by the sequence length for which the percent identity is being reported. “Percent (%) amino acid sequence similarity” is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation. A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids are arginine, lysine and histidine; interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, threonine, cysteine, and glycine.

In some embodiments, a subject protein derivative shares at least 80% sequence identity or similarity, preferably at least 85%, more preferably at least 90%, and most preferably at least 95% sequence identity or similarity with a contiguous stretch of at least 25 amino acids, preferably at least 50 amino acids, more preferably at least 100 amino acids, and in some cases, the entire length of any one of SEQ ID NOS:2, 4, or 6.

The fragment or derivative of a subject protein is preferably “functionally active” meaning that the subject protein derivative or fragment exhibits one or more functional activities associated with a full-length, wild-type subject protein comprising the amino acid sequence of any one of SEQ ID NOS:2, 4, or 6. As one example, a fragment or derivative may have antigenicity such that it can be used in immunoassays, for immunization, for inhibition of activity of a subject protein, etc, as discussed further...
below regarding generation of antibodies to subject proteins. Preferably, a functionally active fragment or derivative of a subject protein is one that displays one or more biological activities associated with a subject protein, such as catalytic activity. For purposes herein, functionally active fragments also include those fragments that exhibit one or more structural features of a subject protein, such as transmembrane or enzymatic domains. The functional activity of the subject proteins, derivatives and fragments can be assayed by various methods known to one skilled in the art (Current Protocols in Protein Science (1998) Coligan et al., eds., John Wiley & Sons, Inc., Somerset, New Jersey). In a preferred method, which is described in detail below, a model organism, such as *Drosophila*, is used in genetic studies to assess the phenotypic effect of a fragment or derivative (i.e. a mutant of a subject protein).

A derivative of a subject protein can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, a cloned subject gene sequence can be cleaved at appropriate sites with restriction endonuclease(s) (Wells et al., Philos. Trans. R. Soc. London Ser A (1986) 317:415), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*, and expressed to produce the desired derivative. Alternatively, a subject gene can be mutated *in vitro or in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or to form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. A variety of mutagenesis techniques are known in the art such as chemical mutagenesis, *in vitro* site-directed mutagenesis (Carter et al., Nucl. Acids Res. (1986) 13:4331), use of TAB\(^*\) linkers (available from Pharmacia and Upjohn, Kalamazoo, MI), etc.

At the protein level, manipulations include post translational modification, e.g. glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known technique (e.g. specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH\(_4\), acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin, etc.). Derivative proteins can also be chemically synthesized by use of a peptide synthesizer, for example to introduce nonclassical amino acids or chemical amino acid analogs as substitutions or additions into the subject protein sequence.

Chimeric or fusion proteins can be made comprising a subject protein or fragment thereof (preferably comprising one or more structural or functional domains of the subject protein) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. Chimeric proteins can be produced by any known method, including: recombinant expression of a nucleic acid encoding the protein (comprising an amino acid sequence encoding a subject protein joined in-frame to a coding sequence for a different protein); ligating the appropriate nucleic acid sequences
encoding the desired amino acid sequences to each other in the proper coding frame, and expressing the
cpheric product; and protein synthetic techniques, e.g. by use of a peptide synthesizer.

Specific subject proteins are discussed below.

**dmLKR/SDH Proteins**

In some embodiments, the invention provides isolated dmLKR/SDH polypeptides. In some of
these embodiments, isolated dmLKR/SDH polypeptides comprise or consist of an amino acid sequence
of SEQ ID NO:2, or fragments or derivatives thereof.

In some embodiments, a dmLKR/SDH polypeptide of the invention comprises an amino acid
sequence of at least about 6, at least about 10, at least about 20, at least about 50, at least about 75, at
least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least
about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about
600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, or
at least about 900 contiguous amino acids of the sequence set forth in SEQ ID NO:2, up to the entire
amino acid sequence as set forth in SEQ ID NO:2.

In one preferred embodiment, a dmLKR/SDH protein derivative shares at least 70% sequence
identity or similarity, preferably at least 80%, more preferably at least 90%, and most preferably at least
95% sequence identity or similarity with a contiguous stretch of at least 25 amino acids, preferably at
least 50 amino acids, more preferably at least 100 amino acids, and in some cases, the entire length of
SEQ ID NO:2.

In another embodiment, the dmLKR/SDH protein derivative may consist of or comprise a
sequence that shares 100% similarity with any contiguous stretch of at least 38 amino acids, preferably
at least 40 amino acids, more preferably at least 43 amino acids, and most preferably at least 48 amino
acids of SEQ ID NO:2, and more preferably of residues 230-835 of SEQ ID NO:2.

Preferred fragments of dmLKR/SDH proteins consist or comprise at least 17, preferably at least
19, more preferably at least 22, and most preferably at least 27 contiguous amino acids of SEQ ID
NO:2, and more preferably of residues 230 to 835 of SEQ ID NO:2.

**dmPAL Proteins**

In some embodiments, the invention provides isolated dmPAL polypeptides. In some of these
embodiments, isolated dmPAL polypeptides comprise or consist of an amino acid sequence of SEQ ID
NO:4, or fragments or derivatives thereof.

In some embodiments, a dmPAL polypeptide of the invention comprises an amino acid sequence
of at least about 6, at least about 10, at least about 20, at least about 50, at least about 75, at least about
100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at
least about 400, at least about 450, at least about 500, or at least about 540 contiguous amino acids of
the sequence set forth in SEQ ID NO:4, up to the entire amino acid sequence as set forth in SEQ ID
NO:4.

In one preferred embodiment, a dmPAL protein derivative shares at least 70% sequence identity
or similarity, preferably at least 80%, more preferably at least 85%, still more preferably at least 90%,
and most preferably at least 95% sequence identity or similarity with a contiguous stretch of at least 25
amino acids, preferably at least 50 amino acids, more preferably at least 100 amino acids, and most
preferably at least 200 amino acids of SEQ ID NO:4.

In another embodiment, the dmPAL protein derivative may consist of or comprise a sequence
that shares 100% similarity with any contiguous stretch of at least 15 amino acids, preferably at least 17
amino acids, more preferably at least 20 amino acids, and most preferably at least 25 amino acids of
SEQ ID NO:4. Preferred derivatives of dmPAL consist of or comprise an amino acid sequence that has
at least 80%, preferably at least 85%, more preferably at least 90%, and most preferably at least 95%
sequence identity or sequence similarity with any of amino acid residues 1-15, 33-463, and 481-541,
which are the likely extracellular or intracellular domains.

Preferred fragments of dmPAL proteins consist or comprise at least 11, preferably at least 13,
more preferably at least 16, and most preferably at least 21 contiguous amino acids of SEQ ID NO:4.

**dmPAL2 Proteins**

In some embodiments, the invention provides isolated dmPAL2 polypeptides. In some of these
embodiments, isolated dmPAL2 polypeptides comprise or consist of an amino acid sequence of SEQ ID
NO:6, or fragments or derivatives thereof.

In some embodiments, a dmPAL2 polypeptide of the invention comprises an amino acid
sequence of at least about 6, at least about 10, at least about 20, at least about 50, at least about 75, at
least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least
about 350, or at least about 400 contiguous amino acids of the sequence set forth in SEQ ID NO:6, up to
the entire amino acid sequence as set forth in SEQ ID NO:6.

In one preferred embodiment, a dmPAL2 protein derivative shares at least 70% sequence
identity or similarity, preferably at least 75%, more preferably at least 80%, still more preferably at least
85%, more preferably at least 90%, and most preferably at least 95% sequence identity or similarity with
a contiguous stretch of at least 25 amino acids, preferably at least 50 amino acids, more preferably at
least 100 amino acids, and in some cases, the entire length of SEQ ID NO:6. Further preferred
derivatives of dmPAL2 consist of or comprise an amino acid sequence that shares the above-listed %
identities or similarities any of amino acid residues 179-206, 230-258, 277-305, and 371-403, which are
putative NHL domains.
In another embodiment, the dmPAL2 protein derivative may consist of or comprise a sequence that shares 100% similarity with any contiguous stretch of at least 18 amino acids, preferably at least 20 amino acids, more preferably at least 23 amino acids, and most preferably at least 28 amino acids of SEQ ID NO:6. Preferred fragments of dmPAL2 proteins consist or comprise at least 12, preferably at least 14, more preferably at least 17, and most preferably at least 22 contiguous amino acids of SEQ ID NO:6.

**GENE REGULATORY ELEMENTS OF THE SUBJECT NUCLEIC ACID MOLECULES**

The invention further provides gene regulatory DNA elements, such as enhancers or promoters that control transcription of the subject nucleic acid molecules. Such regulatory elements can be used to identify tissues, cells, genes and factors that specifically control production of a subject protein. Analyzing components that are specific to a particular subject protein function can lead to an understanding of how to manipulate these regulatory processes, especially for pesticide and therapeutic applications, as well as an understanding of how to diagnose dysfunction in these processes.

Gene fusions with the subject regulatory elements can be made. For compact genes that have relatively few and small intervening sequences, such as those described herein for *Drosophila*, it is typically the case that the regulatory elements that control spatial and temporal expression patterns are found in the DNA immediately upstream of the coding region, extending to the nearest neighboring gene. Regulatory regions can be used to construct gene fusions where the regulatory DNAs are operably fused to a coding region for a reporter protein whose expression is easily detected, and these constructs are introduced as transgenes into the animal of choice. An entire regulatory DNA region can be used, or the regulatory region can be divided into smaller segments to identify sub-elements that might be specific for controlling expression of a given cell type or stage of development. Reporter proteins that can be used for construction of these gene fusions include *E. coli* beta-galactosidase and green fluorescent protein (GFP). These can be detected readily *in situ*, and thus are useful for histological studies and can be used to sort cells that express a subject protein (O’Kane and Gehring PNAS (1987) 84(24):9123-9127; Chalfie *et al.*, Science (1994) 263:802-805; and Cumberledge and Krasnow (1994) Methods in Cell Biology 44:143-159). Recombinase proteins, such as FLP or cre, can be used in controlling gene expression through site-specific recombination (Golic and Lindquist (1989) Cell 59(3):499-509; White *et al.*, Science (1996) 271:805-807). Toxic proteins such as the reaper and hid cell death proteins, are useful to specifically ablate cells that normally express a subject protein in order to assess the physiological function of the cells (Kingston, In Current Protocols in Molecular Biology (1998) Ausubel *et al.*, John Wiley & Sons, Inc. sections 12.0.3-12.10) or any other protein where it is desired to examine the function this particular protein specifically in cells that synthesize a subject protein.
Alternatively, a binary reporter system can be used, similar to that described further below, where a subject regulatory element is operably fused to the coding region of an exogenous transcriptional activator protein, such as the GAL4 or tTA activators described below, to create a subject regulatory element “driver gene”. For the other half of the binary system the exogenous activator controls a separate “target gene” containing a coding region of a reporter protein operably fused to a cognate regulatory element for the exogenous activator protein, such as UAS₂ or a tTA-response element, respectively. An advantage of a binary system is that a single driver gene construct can be used to activate transcription from preconstructed target genes encoding different reporter proteins, each with its own uses as delineated above.

Subject regulatory element-reporter gene fusions are also useful for tests of genetic interactions, where the objective is to identify those genes that have a specific role in controlling the expression of subject genes, or promoting the growth and differentiation of the tissues that expresses a subject protein. Subject gene regulatory DNA elements are also useful in protein-DNA binding assays to identify gene regulatory proteins that control the expression of subject genes. The gene regulatory proteins can be detected using a variety of methods that probe specific protein-DNA interactions well known to those skilled in the art (Kingston, supra) including in vivo footprinting assays based on protection of DNA sequences from chemical and enzymatic modification within living or permeabilized cells; and in vitro footprinting assays based on protection of DNA sequences from chemical or enzymatic modification using protein extracts, nitrocellulose filter-binding assays and gel electrophoresis mobility shift assays using radioactively labeled regulatory DNA elements mixed with protein extracts. Candidate gene regulatory proteins can be purified using a combination of conventional and DNA-affinity purification techniques. Molecular cloning strategies can also be used to identify proteins that specifically bind subject gene regulatory DNA elements. For example, a Drosophila cDNA library in an expression vector, can be screened for cDNAs that encode subject gene regulatory element DNA-binding activity. Similarly, the yeast “one-hybrid” system can be used (Li and Herskowitz, Science (1993) 262:1870-1874; Luo et al., Biotechniques (1996) 20(4):564-568; Vidal et al., PNAS (1996) 93(19):10315-10320).

**dmLKR/SDH Gene Regulatory Elements**

In some embodiments, dmLKR/SDH regulatory elements are provided. dmLKR/SDH gene regulatory DNA elements, such as enhancers or promoters that reside within nucleotides 1 to 103, can be used to identify tissues, cells, genes and factors that specifically control dmLKR/SDH protein production. Preferably at least 20, more preferably at least 25, and most preferably at least 50 contiguous nucleotides within nucleotides 1 to 103 of SEQ ID NO:1 are used. Analyzing components that are specific to dmLKR/SDH protein function can lead to an understanding of how to manipulate
these regulatory processes, especially for pesticide and therapeutic applications, as well as an understanding of how to diagnose dysfunction in these processes.

**dmPAL Gene Regulatory Elements**

dmPAL gene regulatory DNA elements, such as enhancers or promoters that reside within nucleotides 1 to 521, can be used to identify tissues, cells, genes and factors that specifically control dmPAL protein production. Preferably at least 50, more preferably at least 100, and most preferably at least 150 contiguous nucleotides within nucleotides 1 to 521 of SEQ ID NO:3 are used. Analyzing components that are specific to dmPAL protein function can lead to an understanding of how to manipulate these regulatory processes, especially for pesticide and therapeutic applications, as well as an understanding of how to diagnose dysfunction in these processes.

**dmPAL2 Gene Regulatory Elements**

dmPAL2 gene regulatory DNA elements, such as enhancers or promoters that reside within nucleotides 1 to 112, can be used to identify tissues, cells, genes and factors that specifically control dmPAL2 protein production. Preferably at least 20, more preferably at least 25, and most preferably at least 50 contiguous nucleotides within nucleotides 1 to 112 of SEQ ID NO:5 are used. Analyzing components that are specific to dmPAL2 protein function can lead to an understanding of how to manipulate these regulatory processes, especially for pesticide and therapeutic applications, as well as an understanding of how to diagnose dysfunction in these processes.

**Antibodies Specific for Subject Proteins**

The present invention provides antibodies, which may be isolated antibodies, that bind specifically to a subject protein. The subject proteins, fragments thereof, and derivatives thereof may be used as an immunogen to generate monoclonal or polyclonal antibodies and antibody fragments or derivatives (e.g. chimeric, single chain, Fab fragments). For example, fragments of a subject protein, preferably those identified as hydrophilic, are used as immunogens for antibody production using art-known methods such as by hybridomas; production of monoclonal antibodies in germ-free animals (PCT/US90/02545); the use of human hybridomas (Cole et al., PNAS (1983) 80:2026-2030; Cole et al., in Monoclonal Antibodies and Cancer Therapy (1985) Alan R. Liss, pp. 77-96), and production of humanized antibodies (Jones et al., Nature (1986) 321:522-525; U.S. Pat. 5,530,101). In a particular embodiment, subject polypeptide fragments provide specific antigens and/or immunogens, especially when coupled to carrier proteins. For example, peptides are covalently coupled to keyhole limpet antigen (KLH) and the conjugate is emulsified in Freund's complete adjuvant. Laboratory rabbits are immunized according to conventional protocol and bled. The presence of specific antibodies is assayed
by solid phase immunosorbent assays using immobilized corresponding polypeptide. Specific activity or function of the antibodies produced may be determined by convenient in vitro, cell-based, or in vivo assays: e.g. in vitro binding assays, etc. Binding affinity may be assayed by determination of equilibrium constants of antigen-antibody association (usually at least about $10^7 \text{M}^{-1}$, preferably at least about $10^8 \text{M}^{-1}$, more preferably at least about $10^9 \text{M}^{-1}$).

### Identification of Molecules that Interact with a Subject Protein

A variety of methods can be used to identify or screen for molecules, such as proteins or other molecules, that interact with a subject protein, or derivatives or fragments thereof. The assays may employ purified protein, or cell lines or model organisms such as *Drosophila* and *C. elegans*, that have been genetically engineered to express a subject protein. Suitable screening methodologies are well known in the art to test for proteins and other molecules that interact with a subject gene and protein (see e.g., PCT International Publication No. WO 96/34099). The newly identified interacting molecules may provide new targets for pharmaceutical or pesticidal agents. Any of a variety of exogenous molecules, both naturally occurring and/or synthetic (e.g., libraries of small molecules or peptides, or phage display libraries), may be screened for binding capacity. In a typical binding experiment, a subject protein or fragment is mixed with candidate molecules under conditions conducive to binding, sufficient time is allowed for any binding to occur, and assays are performed to test for bound complexes. Assays to find interacting proteins can be performed by any method known in the art, for example, immunoprecipitation with an antibody that binds to the protein in a complex followed by analysis by size fractionation of the immunoprecipitated proteins (e.g. by denaturing or non-denaturing polyacrylamide gel electrophoresis), Western analysis, non-denaturing gel electrophoresis, two-hybrid systems (Fields and Song, Nature (1989) 340:245-246; U.S. Pat. NO. 5,283,173; for review see Brent and Finley, Annu. Rev. Genet. (1977) 31:663-704), etc.

### Immunoassays

Immunoassays can be used to identify proteins that interact with or bind to a subject protein. Various assays are available for testing the ability of a protein to bind to or compete with binding to a wild-type subject protein or for binding to an anti-subject protein antibody. Suitable assays include radioimmunoassays, ELISA (enzyme linked immunosorbent assay), immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, in situ immunoassays (e.g., using colloidal gold, enzyme or radioisotope labels), western blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, immunoelectrophoresis assays, etc.
Identification of Potential Pesticide or Drug Targets

Once new target genes or target interacting genes are identified, they can be assessed as potential pesticide or drug targets, or as potential biopesticides. Further, transgenic plants that express subject proteins can be tested for activity against insect pests (Estruch et al., Nat. Biotechnol (1997) 15(2):137-141).

The subject proteins are validated pesticide targets, since disruption of the Drosophila the subject genes results in lethality when homozygous. The mutation to lethality of these gene indicates that drugs that agonize or antagonize the gene product may be effective pesticidal agents.

As used herein, the term “pesticide” refers generally to chemicals, biological agents, and other compounds that kill, paralyze, sterilize or otherwise disable pest species in the areas of agricultural crop protection, human and animal health. Exemplary pest species include parasites and disease vectors such as mosquitoes, fleas, ticks, parasitic nematodes, chiggers, mites, etc. Pest species also include those that are eradicated for aesthetic and hygienic purposes (e.g. ants, cockroaches, clothes moths, flour beetles, etc.), home and garden applications, and protection of structures (including wood boring pests such as termites, and marine surface fouling organisms).

Pesticidial compounds can include traditional small organic molecule pesticides (typified by compound classes such as the organophosphates, pyrethroids, carbamates, and organochlorines, benzoyleureas, etc.). Other pesticides include proteinaceous toxins such as the Bacillus thuringiensis Crytoxins (Gill et al., Annu Rev Entomol (1992) 37:615-636) and Photorhabdus luminescens toxins (Bowden et al., Science (1998) 280:2129-2132); and nucleic acids such as subject dsRNA or antisense nucleic acids that interfere with activity of a subject nucleic acid molecule. Pesticides can be delivered by a variety of means including direct application to pests or to their food source. In addition to direct application, toxic proteins and pesticidal nucleic acids (e.g. dsRNA) can be administered using biopesticidal methods, for example, by viral infection with nucleic acid or by transgenic plants that have been engineered to produce interfering nucleic acid sequences or encode the toxic protein, which are ingested by plant-eating pests.

Putative pesticides, drugs, and molecules can be applied onto whole insects, nematodes, and other small invertebrate metazoans, and the ability of the compounds to modulate (e.g. block or enhance) activity of a subject protein can be observed. Alternatively, the effect of various compounds on a subject protein can be assayed using cells that have been engineered to express one or more subject proteins and associated proteins.

Assays of Compounds on Worms

In a typical worm assay, the compounds to be tested are dissolved in DMSO or other organic solvent, mixed with a bacterial suspension at various test concentrations, preferably OP50 strain of
bacteria (Brenner, Genetics (1974) 110:421-440), and supplied as food to the worms. The population of
worms to be treated can be synchronized larvae (Sulston and Hodgkin, in the nematode C. elegans
(1988), supra) or adults or a mixed-stage population of animals.

Adult and larval worms are treated with different concentrations of compounds, typically
ranging from 1 mg/ml to 0.001 mg/ml. Behavioral aberrations, such as a decrease in motility and
growth, and morphological aberrations, sterility, and death are examined in both acutely and chronically
treated adult and larval worms. For the acute assay, larval and adult worms are examined immediately
after application of the compound and re-examined periodically (every 30 minutes) for 5-6 hours.
Chronic or long-term assays are performed on worms and the behavior of the treated worms is examined
every 8-12 hours for 4-5 days. In some circumstances, it is necessary to reapply the pesticide to the
treated worms every 24 hours for maximal effect.

Assays of Compounds on Insects

Potential insecticidal compounds can be administered to insects in a variety of ways, including
orally (including addition to synthetic diet, application to plants or prey to be consumed by the test
organism), topically (including spraying, direct application of compound to animal, allowing animal to
contact a treated surface), or by injection. Insecticides are typically very hydrophobic molecules and
must commonly be dissolved in organic solvents, which are allowed to evaporate in the case of methanol
or acetone, or at low concentrations can be included to facilitate uptake (ethanol, dimethyl sulfoxide).

The first step in an insect assay is usually the determination of the minimal lethal dose (MLD)
on the insects after a chronic exposure to the compounds. The compounds are usually diluted in DMSO,
and applied to the food surface bearing 0-48 hour old embryos and larvae. In addition to MLD, this step
allows the determination of the fraction of eggs that hatch, behavior of the larvae, such as how they move
/feed compared to untreated larvae, the fraction that survive to pupate, and the fraction that eclose
(emergence of the adult insect from puparium). Based on these results more detailed assays with shorter
exposure times may be designed, and larvae might be dissected to look for obvious morphological
defects. Once the MLD is determined, more specific acute and chronic assays can be designed.

In a typical acute assay, compounds are applied to the food surface for embryos, larvae, or
adults, and the animals are observed after 2 hours and after an overnight incubation. For application on
embryos, defects in development and the percent that survive to adulthood are determined. For larvae,
defects in behavior, locomotion, and molting may be observed. For application on adults, defects in
levels and/or enzyme activity are observed, and effects on behavior and/or fertility are noted.

For a chronic exposure assay, adults are placed on vials containing the compounds for 48 hours,
then transferred to a clean container and observed for fertility, defects in levels and/or activity of a
subject enzyme, and death.
Assay of Compounds using Cell Cultures

Compounds that modulate (e.g. block or enhance) a subject protein's activity may also be assayed using cell culture. For example, various compounds added to cells expressing a subject protein may be screened for their ability to modulate the activity of subject genes based upon measurements of a biological activity of a subject protein. For example, compounds may be screened for their ability to modulate the activity of dmLKR/SDH genes based on measurements of reductase or dehydrogenase activity. Compounds may be screened for their ability to modulate dmPAL genes or dmPAL genes based on measurements of enzyme activity. Assays for changes in a biological activity of a subject protein can be performed on cultured cells expressing endogenous normal or mutant subject protein.

Such studies also can be performed on cells transfected with vectors capable of expressing the subject protein, or functional domains of one of the subject protein, in normal or mutant form. In addition, to enhance the signal measured in such assays, cells may be cotransfected with genes encoding a subject protein.

Alternatively, cells expressing a subject protein may be lysed, the subject protein purified, and tested in vitro using methods known in the art (Kanemaki M., et al., J Biol Chem, 1999 274:22437-22444).

Compounds that selectively modulate a subject protein are identified as potential pesticide and drug candidates having specificity for the subject protein.

Identification of small molecules and compounds as potential pesticides or pharmaceutical compounds from large chemical libraries requires high-throughput screening (HTS) methods (Bolger, Drug Discovery Today (1999) 4:251-253). Several of the assays mentioned herein can lend themselves to such screening methods. For example, cells or cell lines expressing wild type or mutant subject protein or its fragments, and a reporter gene can be subjected to compounds of interest, and depending on the reporter genes, interactions can be measured using a variety of methods such as color detection, fluorescence detection (e.g. GFP), autoradiography, scintillation analysis, etc.

Compounds identified using the above-described methods are useful to control pests, e.g., are useful as pesticides. Such compounds can control pests, e.g., by reducing pest growth, and/or fertility, and/or viability.

Subject Nucleic Acids as Biopesticides

Subject nucleic acids and fragments thereof, such as antisense sequences or double-stranded RNA (dsRNA), can be used to inhibit subject nucleic acid molecule function, and thus can be used as biopesticides. Methods of using dsRNA interference are described in published PCT application WO 99/32619. The biopesticides may comprise the nucleic acid molecule itself, an expression construct capable of expressing the nucleic acid, or organisms transfected with the expression construct. The
biopesticides may be applied directly to plant parts or to soil surrounding the plants (e.g. to access plant parts growing beneath ground level), or directly onto the pest.


**Generation and Genetic Analysis of Animals and Cell Lines with Altered Expression of a Subject Gene**

Both genetically modified animal models (i.e. *in vivo* models), such as *C. elegans* and *Drosophila*, and *in vitro* models such as genetically engineered cell lines expressing or mis-expressing subject pathway genes, are useful for the functional analysis of these proteins. Model systems that display detectable phenotypes, can be used for the identification and characterization of subject pathway genes or other genes of interest and/or phenotypes associated with the mutation or mis-expression of subject pathway protein. The term “mis-expression” as used herein encompasses mis-expression due to
gene mutations. Thus, a mis-expressed subject pathway protein may be one having an amino acid sequence that differs from wild-type (i.e. it is a derivative of the normal protein). A mis-expressed subject pathway protein may also be one in which one or more amino acids have been deleted, and thus is a "fragment" of the normal protein. As used herein, "mis-expression" also includes ectopic expression (e.g. by altering the normal spatial or temporal expression), over-expression (e.g. by multiple gene copies), underexpression, non-expression (e.g. by gene knockout or blocking expression that would otherwise normally occur), and further, expression in ectopic tissues. As used in the following discussion concerning in vivo and in vitro models, the term "gene of interest" refers to a subject pathway gene, or any other gene involved in regulation or modulation, or downstream effector of the subject pathway.

The in vivo and in vitro models may be genetically engineered or modified so that they 1) have deletions and/or insertions of one or more subject pathway genes, 2) harbor interfering RNA sequences derived from subject pathway genes, 3) have had one or more endogenous subject pathway genes mutated (e.g. contain deletions, insertions, rearrangements, or point mutations in subject gene or other genes in the pathway), and/or 4) contain transgenes for mis-expression of wild-type or mutant forms of such genes. Such genetically modified in vivo and in vitro models are useful for identification of genes and proteins that are involved in the synthesis, activation, control, etc. of subject pathway gene and/or gene products, and also downstream effectors of subject function, genes regulated by subject, etc. The newly identified genes could constitute possible pesticide targets (as judged by animal model phenotypes such as non-viability, block of normal development, defective feeding, defective movement, or defective reproduction). The model systems can also be used for testing potential pesticidal or pharmaceutical compounds that interact with the subject pathway, for example by administering the compound to the model system using any suitable method (e.g. direct contact, ingestion, injection, etc.) and observing any changes in phenotype, for example defective movement, lethality, etc. Various genetic engineering and expression modification methods which can be used are well-known in the art, including chemical mutagenesis, transposon mutagenesis, antisense RNAi, dsRNAi, and transgene-mediated mis-expression.

Generating Loss-of-function Mutations by Mutagenesis

use of radiation (e.g., X-ray, UV, or gamma ray); chemicals (e.g., EMS, MMS, ENU, formaldehyde, etc.); and insertional mutagenesis by mobile elements including dysgenesis induced by transposon insertions, or transposon-mediated deletions, for example, male recombination, as described below. Other methods of altering expression of genes include use of transposons (e.g., P element, EP-type “overexpression trap” element, mariner element, piggyBac transposon, hermes, minos, sleeping beauty, etc.) to misexpress genes; antisense; double-stranded RNA interference; peptide and RNA aptamers; directed deletions; homologous recombination; dominant negative alleles; and intrabodies.

Transposon insertions lying adjacent to a gene of interest can be used to generate deletions of flanking genomic DNA, which if induced in the germline, are stably propagated in subsequent generations. The utility of this technique in generating deletions has been demonstrated and is well-known in the art. One version of the technique using collections of P element transposon induced recessive lethal mutations (P lethals) is particularly suitable for rapid identification of novel, essential genes in Drosophila (Cooley et al., Science (1988) 239:1121-1128; Spralding et al., PNAS (1995) 92:0824-10830). Since the sequence of the P elements are known, the genomic sequence flanking each transposon insert is determined either by plasmid rescue (Hamilton et al., PNAS (1991) 88:2731-2735) or by inverse polymerase chain reaction, using well-established techniques. (Rehm, http://www.fruitfly.org/methods/). The subject genes were identified from a P lethal screen. Disruption of the Drosophila subject gene results in lethality when homozygous, indicating that this protein is critical for cell function and the survival of insects. The mutation to lethality of this gene indicates that drugs which agonize or antagonize the encoded subject protein will be effective insecticidal agents and that this class of proteins are excellent targets for drug screening and discovery.

A more recent version of the transposon insertion technique in male Drosophila using P elements is known as P-mediated male recombination (Preston and Engels, Genetics (1996) 144:1611-1638).

**Generating Loss-of-function Phenotypes Using RNA-based Methods**

The subject genes may be identified and/or characterized by generating loss-of-function phenotypes in animals of interest through RNA-based methods, such as antisense RNA (Schubiger and Edgar, Methods in Cell Biology (1994) 44:697-713). One form of the antisense RNA method involves the injection of embryos with an antisense RNA that is partially homologous to the gene of interest (in this case the subject gene). Another form of the antisense RNA method involves expression of an antisense RNA partially homologous to the gene of interest by operably joining a portion of the gene of interest in the antisense orientation to a powerful promoter that can drive the expression of large quantities of antisense RNA, either generally throughout the animal or in specific tissues. Antisense RNA-generated loss-of-function phenotypes have been reported previously for several Drosophila genes.


Another method for generating loss-of-function phenotypes is by double-stranded RNA interference (dsRNAi). This method is based on the interfering properties of double-stranded RNA derived from the coding regions of gene, and has proven to be of great utility in genetic studies of C. elegans (Fire et al., Nature (1998) 391:806-811), and can also be used to generate loss-of-function phenotypes in Drosophila (Kennerdell and Carthew, Cell (1998) 95:1017-1026; Misquitta and Patterson PNAS (1999) 96:1451-1456). In one example of this method, complementary sense and antisense RNAs derived from a substantial portion of a gene of interest, such as a subject gene, are synthesized in vitro. The resulting sense and antisense RNAs are annealed in an injection buffer, and the double-stranded RNA injected or otherwise introduced into animals (such as in their food or by soaking in the buffer containing the RNA). Progeny of the injected animals are then inspected for phenotypes of interest (PCT publication no. WO99/32619).

In another embodiment of the method, the dsRNA can be delivered to the animal by bathing the animal in a solution containing a sufficient concentration of the dsRNA. In another embodiment of the method, dsRNA derived from the subject genes can be generated in vivo by simultaneous expression of both sense and antisense RNA from appropriately positioned promoters operably fused to subject sequences in both sense and antisense orientations. In yet another embodiment of the method the dsRNA can be delivered to the animal by engineering expression of dsRNA within cells of a second organism that serves as food for the animal, for example engineering expression of dsRNA in E. coli bacteria which are fed to C. elegans, or engineering expression of dsRNA in baker’s yeast which are fed to Drosophila, or engineering expression of dsRNA in transgenic plants which are fed to plant eating insects such as Leptinotarsa or Heliiothis. Recently, RNAi has been successfully used in cultured Drosophila cells to inhibit expression of targeted proteins (Clemens, J.C., et al., Proc Natl Acad Sci U S A 2000 Jun 6;97(12):6499-503). Thus, cell lines in culture can be manipulated using RNAi both to perturb and study the function of the subject gene pathway components and to validate the efficacy of therapeutic or pesticidal strategies that involve the manipulation of this pathway.
Generating Loss-of-function Phenotypes Using Peptide and RNA Aptamers


Generating Loss of Function Phenotypes Using Intrabodies

Intracellularly expressed antibodies, or intrabodies, are single-chain antibody molecules designed to specifically bind and inactivate target molecules inside cells. Intrabodies have been used in cell assays and in whole organisms such as Drosophila (Chen et al., Hum. Gen. Ther. (1994) 5:595-601; Hassanzadeh et al., Febs Lett. (1998) 16(1, 2):75-80 and 81-86). Inducible expression vectors can be constructed with intrabodies that react specifically with a subject protein. These vectors can be introduced into model organisms and studied in the same manner as described above for aptamers.

Transgenesis

Typically, transgenic animals are created that contain gene fusions of the coding regions of a subject gene (from either genomic DNA or cDNA) or genes engineered to encode antisense RNAs, cosuppression RNAs, interfering dsRNA, RNA aptamers, peptide aptamers, or intrabodies operably joined to a specific promoter and transcriptional enhancer whose regulation has been well characterized, preferably heterologous promoters/enhancers (i.e. promoters/enhancers that are non-native to a subject pathway genes being expressed).

Methods are well known for incorporating exogenous nucleic acid sequences into the genome of animals or cultured cells to create transgenic animals or recombinant cell lines. For invertebrate animal models, the most common methods involve the use of transposable elements. There are several suitable transposable elements that can be used to incorporate nucleic acid sequences into the genome of model organisms. Transposable elements are particularly useful for inserting sequences into a gene of interest so that the encoded protein is not properly expressed, creating a “knock-out” animal having a loss-of-function phenotype. Techniques are well-established for the use of P element in Drosophila (Rubin and Spradling, Science (1982) 218:348-53; U.S. Pat. No. 4,670,388) and Tc1 in C. elegans (Zwaal et al., Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7431-7435; and Caenorhabditis elegans: Modern Biological Analysis of an Organism (1995) Epstein and Shakes, Eds.). Other Tc1-like transposable elements can
be used such as minos, mariner and sleeping beauty. Additionally, transposable elements that function in a variety of species, have been identified, such as PiggyBac (Thibault et al., Insect Mol Biol (1999) 8(1):119-23), hobo, and hermes.

P elements, or marked P elements, are preferred for the isolation of loss-of-function mutations in *Drosophila* genes because of the precise molecular mapping of these genes, depending on the availability and proximity of preexisting P element insertions for use as a localized transposon source (Hamilton and Zinn, Methods in Cell Biology (1994) 44:81-94; and Wolfner and Goldberg, Methods in Cell Biology (1994) 44:33-80). Typically, modified P elements are used which contain one or more elements that allow detection of animals containing the P element. Most often, marker genes are used that affect the eye color of *Drosophila*, such as derivatives of the *Drosophila white* or *rosy* genes (Rubin and Spradling, Science (1982) 218(4570):348-353; and Klemenz et al., Nucleic Acids Res. (1987) 15(10):3947-3959). However, in principle, any gene can be used as a marker that causes a reliable and easily scored phenotypic change in transgenic animals. Various other markers include bacterial plasmid sequences having selectable markers such as ampicillin resistance (Steller and Pirrotta, EMBO. J. (1985) 4:167-171); and lacZ sequences fused to a weak general promoter to detect the presence of enhancers with a developmental expression pattern of interest (Bellen et al., Genes Dev. (1989) 3(9):1288-1300). Other examples of marked P elements useful for mutagenesis have been reported (Nucleic Acids Research (1998) 26:85-88; and http://flybase.bio.indiana.edu).


In addition to creating loss-of-function phenotypes, transposable elements can be used to incorporate the gene of interest, or mutant or derivative thereof, as an additional gene into any region of an animal’s genome resulting in mis-expression (including over-expression) of the gene. A preferred vector designed specifically for misexpression of genes in transgenic *Drosophila*, is derived from pGMR (Hay et al., Development (1994) 120:2121-2129), is 9Kb long, and contains: an origin of replication for *E. coli*; an ampicillin resistance gene; P element transposon 3' and 5' ends to mobilize the inserted sequences; a White marker gene; an expression unit comprising the TATA region of hsp70 enhancer and the 3' untranslated region of α-tubulin gene. The expression unit contains a first multiple cloning site (MCS) designed for insertion of an enhancer and a second MCS located 500 bases downstream, designed for the insertion of a gene of interest. As an alternative to transposable elements, homologous
recombination or gene targeting techniques can be used to substitute a gene of interest for one or both copies of the animal’s homologous gene. The transgene can be under the regulation of either an exogenous or an endogenous promoter element, and be inserted as either a minigene or a large genomic fragment. In one application, gene function can be analyzed by ectopic expression, using, for example, *Drosophila* (Brand et al., Methods in Cell Biology (1994) 44:635-654) or *C. elegans* (Mello and Fire, Methods in Cell Biology (1995) 48:451-482).

Examples of well-characterized heterologous promoters that may be used to create the transgenic animals include heat shock promoters/enhancers, which are useful for temperature induced mis-expression. In *Drosophila*, these include the *hsp70* and *hsp83* genes, and in *C. elegans*, include *hsp16-2* and *hsp16-41*. Tissue specific promoters/enhancers are also useful, and in *Drosophila*, include *eyeless* (Mozer and Benzer, Development (1994) 120:1049-1058), *sevenless* (Bowtell et al., PNAS (1991) 88(15):6853-6857), and *glass*-responsive promoters/enhancers (Quiring et al., Science (1994) 265:785-789) which are useful for expression in the eye; and enhancers/promoters derived from the *dpp* or *vestigial* genes which are useful for expression in the wing (Staehling-Hampton et al., Cell Growth Differ. (1994) 5(6):585-593; Kim et al., Nature (1996) 382:133-138). Finally, where it is necessary to restrict the activity of dominant active or dominant negative transgenes to regions where the pathway is normally active, it may be useful to use endogenous promoters of genes in the pathway, such as a subject protein pathway genes.

In *C. elegans*, examples of useful tissue specific promoters/enhancers include the *myo-2* gene promoter, useful for pharyngeal muscle-specific expression; the *hli-1* gene promoter, useful for body-muscle-specific expression; and the gene promoter, useful for touch-neuron-specific gene expression. In a preferred embodiment, gene fusions for directing the mis-expression of a subject pathway gene are incorporated into a transformation vector which is injected into nematodes along with a plasmid containing a dominant selectable marker, such as *rol-6*. Transgenic animals are identified as those exhibiting a roller phenotype, and the transgenic animals are inspected for additional phenotypes of interest created by mis-expression of a subject pathway gene.


Dominant negative mutations, by which the mutation causes a protein to interfere with the normal function of a wild-type copy of the protein, and which can result in loss-of-function or reduced-

Assays for Change in Gene Expression

Various expression analysis techniques may be used to identify genes which are differentially expressed between a cell line or an animal expressing a wild type subject gene compared to another cell line or animal expressing a mutant subject gene. Such expression profiling techniques include differential display, serial analysis of gene expression (SAGE), transcript profiling coupled to a gene database query, nucleic acid array technology, subtractive hybridization, and proteome analysis (e.g. mass-spectrometry and two-dimensional protein gels). Nucleic acid array technology may be used to determine a global (i.e., genome-wide) gene expression pattern in a normal animal for comparison with an animal having a mutation in a subject gene. Gene expression profiling can also be used to identify other genes (or proteins) that may have a functional relation to a subject (e.g. may participate in a signaling pathway with a subject gene). The genes are identified by detecting changes in their expression levels following mutation, i.e., insertion, deletion or substitution in, or over-expression, under-expression, mis-expression or knock-out, of the subject.

Phenotypes Associated with Target Pathway Gene Mutations

After isolation of model animals carrying mutated or mis-expressed subject pathway genes or inhibitory RNAs, animals are carefully examined for phenotypes of interest. For analysis of subject pathway genes that have been mutated (i.e. deletions, insertions, and/or point mutations) animal models that are both homozygous and heterozygous for the altered subject pathway gene are analyzed. Examples of specific phenotypes that may be investigated include lethality; sterility; feeding behavior, perturbations in neuromuscular function including alterations in motility, and alterations in sensitivity to pesticides and pharmaceuticals. Some phenotypes more specific to flies include alterations in: adult behavior such as, flight ability, walking, grooming, phototaxis, mating or egg-laying; alterations in the responses of sensory organs, changes in the morphology, size or number of adult tissues such as, eyes, wings, legs, bristles, antennae, gut, fat body, gonads, and musculature; larval tissues such as mouth parts, cuticles, internal tissues or imaginal dises; or larval behavior such as feeding, molting, crawling, or puparian formation; or developmental defects in any germline or embryonic tissues. Some phenotypes more specific to nematodes include: locomotory, egg laying, chemosensation, male mating, and intestinal expulsion defects. In various cases, single phenotypes or a combination of specific phenotypes in model organisms might point to specific genes or a specific pathway of genes, which facilitate the cloning process.
Genomic sequences containing a subject pathway gene can be used to confirm whether an existing mutant insect or worm line corresponds to a mutation in one or more subject pathway genes, by rescuing the mutant phenotype. Briefly, a genomic fragment containing the subject pathway gene of interest and potential flanking regulatory regions can be subcloned into any appropriate insect (such as Drosophila) or worm (such as C. elegans) transformation vector, and injected into the animals. For Drosophila, an appropriate helper plasmid is used in the injections to supply transposase for transposon-based vectors. Resulting germine transformants are crossed for complementation testing to an existing or newly created panel of Drosophila or C. elegans lines whose mutations have been mapped to the vicinity of the gene of interest (Fly Pushing: The Theory and Practice of Drosophila Genetics, supra; and Caenorhabditis elegans: Modern Biological Analysis of an Organism (1995), Epstein and Shakes, eds.). If a mutant line is discovered to be rescued by this genomic fragment, as judged by complementation of the mutant phenotype, then the mutant line likely harbors a mutation in the subject pathway gene. This prediction can be further confirmed by sequencing the subject pathway gene from the mutant line to identify the lesion in the subject pathway gene.

Identification of Genes That Modify a Subject Genes

The characterization of new phenotypes created by mutations or misexpression in subject genes enables one to test for genetic interactions between subject genes and other genes that may participate in the same, related, or interacting genetic or biochemical pathway(s). Individual genes can be used as starting points in large-scale genetic modifier screens as described in more detail below. Alternatively, RNAi methods can be used to simulate loss-of-function mutations in the genes being analyzed. It is of particular interest to investigate whether there are any interactions of subject genes with other well-characterized genes, particularly genes involved in DNA unwinding.

Genetic Modifier Screens

A genetic modifier screen using invertebrate model organisms is a particularly preferred method for identifying genes that interact with subject genes, because large numbers of animals can be systematically screened making it more possible that interacting genes will be identified. In Drosophila, a screen of up to about 10,000 animals is considered to be a pilot-scale screen. Moderate-scale screens usually employ about 10,000 to about 50,000 flies, and large-scale screens employ greater than about 50,000 flies. In a genetic modifier screen, animals having a mutant phenotype due to a mutation in or misexpression of one or more subject genes are further mutagenized, for example by chemical mutagenesis or transposon mutagenesis.

The procedures involved in typical Drosophila genetic modifier screens are well known in the art (Wolfner and Goldberg, Methods in Cell Biology (1994) 44:33-80; and Karim et al., Genetics (1996)).
The procedures used differ depending upon the precise nature of the mutant allele being modified. If the mutant allele is genetically recessive, as is commonly the situation for a loss-of-function allele, then most typically males, or in some cases females, which carry one copy of the mutant allele are exposed to an effective mutagen, such as EMS, MMS, ENU, triethylamine, diepoxyalkanes, ICR-170, formaldehyde, X-rays, gamma rays, or ultraviolet radiation. The mutagenized animals are crossed to animals of the opposite sex that also carry the mutant allele to be modified. In the case where the mutant allele being modified is genetically dominant, as is commonly the situation for ectopically expressed genes, wild type males are mutagenized and crossed to females carrying the mutant allele to be modified.

The progeny of the mutagenized and crossed flies that exhibit either enhancement or suppression of the original phenotype are presumed to have mutations in other genes, called “modifier genes”, that participate in the same phenotype-generating pathway. These progeny are immediately crossed to adults containing balancer chromosomes and used as founders of a stable genetic line. In addition, progeny of the founder adult are retested under the original screening conditions to ensure stability and reproducibility of the phenotype. Additional secondary screens may be employed, as appropriate, to confirm the suitability of each new modifier mutant line for further analysis.

Standard techniques used for the mapping of modifiers that come from a genetic screen in Drosophila include meiotic mapping with visible or molecular genetic markers; male-specific recombination mapping relative to P-element insertions; complementation analysis with deficiencies, duplications, and lethal P-element insertions; and cytological analysis of chromosomal aberrations (Fly Pushing: Theory and Practice of Drosophila Genetics, supra; Drosophila: A Laboratory Handbook, supra). Genes corresponding to modifier mutations that fail to complement a lethal P-element may be cloned by plasmid rescue of the genomic sequence surrounding that P-element. Alternatively, modifier genes may be mapped by phenotype rescue and positional cloning (Sambrook et al., supra).

Newly identified modifier mutations can be tested directly for interaction with other genes of interest known to be involved or implicated with a subject gene using methods described above. Also, the new modifier mutations can be tested for interactions with genes in other pathways that are not believed to be related to neuronal signaling (e.g. nanos in Drosophila). New modifier mutations that exhibit specific genetic interactions with other genes implicated in neuronal signaling, but not interactions with genes in unrelated pathways, are of particular interest.

The modifier mutations may also be used to identify “complementation groups”. Two modifier mutations are considered to fall within the same complementation group if animals carrying both mutations in trans exhibit essentially the same phenotype as animals that are homozygous for each mutation individually and, generally are lethal when in trans to each other (Fly Pushing: The Theory and Practice of Drosophila Genetics, supra). Generally, individual complementation groups defined in this way correspond to individual genes.
When modifier genes are identified, homologous genes in other species can be isolated using procedures based on cross-hybridization with modifier gene DNA probes, PCR-based strategies with primer sequences derived from the modifier genes, and/or computer searches of sequence databases. For therapeutic applications related to the function of subject genes, human and rodent homologs of the modifier genes are of particular interest. For pesticide and other agricultural applications, homologs of modifier genes in insects and arachnids are of particular interest. Insects, arachnids, and other organisms of interest include, among others, Isopoda; Diplopoda; Chilopoda; Symphyla; Thysanura; Collembola; Orthoptera, such as *Scistocerca* spp.; Blattoidea, such as *Blattella germanica*; Dermaptera; Isoptera; Anoplura; Mallophaga; Thysanoptera; Heteroptera; Homoptera, including *Bemisia* tabaci, and *Myzus* spp.; Lepidoptera including *Plodia interpunctella, Pectinophora gossypiella, Plutella* spp., *Heliothis* spp., and *Spodoptera* species; Coleoptera such as *Lepinotarsa, Diabrotica* spp., *Anthonomus* spp., and *Tribolium* spp.; Hymenoptera; Diptera, including *Anopheles* spp.; Siphonaptera, including Ctenocephalides felis; Arachnida; and Aracnids, including *Amblyomma americanum*; and nematodes, including *Meloidogyne* spp., and *Heterodera glycines*.

Although the above-described *Drosophila* genetic modifier screens are quite powerful and sensitive, some genes that interact with subject genes may be missed in this approach, particularly if there is functional redundancy of those genes. This is because the vast majority of the mutations generated in the standard mutagenesis methods will be loss-of-function mutations, whereas gain-of-function mutations that could reveal genes with functional redundancy will be relatively rare. Another method of genetic screening in *Drosophila* has been developed that focuses specifically on systematic gain-of-function genetic screens (Rorth et al., Development (1998) 125:1049-1057). This method is based on a modular mis-expression system utilizing components of the GAL4/UAS system (described above) where a modified P element, termed an "enhanced P" (EP) element, is genetically engineered to contain a GAL4-responsive UAS element and promoter. Any other transposons can also be used for this system. The resulting transposon is used to randomly tag genes by insertional mutagenesis (similar to the method of P element mutagenesis described above). Thousands of transgenic *Drosophila* strains, termed EP lines, can be generated, each containing a specific UAS-tagged gene. This approach takes advantage of the preference of P elements to insert at the 5'-ends of genes. Consequently, many of the genes that are tagged by insertion of EP elements become operably fused to a GAL4-regulated promoter, and increased expression or mis-expression of the randomly tagged gene can be induced by crossing in a GAL4 driver gene.

Systematic gain-of-function genetic screens for modifiers of phenotypes induced by mutation or mis-expression of a subject gene can be performed by crossing several thousand *Drosophila* EP lines individually into a genetic background containing a mutant or mis-expressed subject gene, and further containing an appropriate GAL4 driver transgene. It is also possible to remobilize the EP elements to
obtain novel insertions. The progeny of these crosses are then analyzed for enhancement or suppression of the original mutant phenotype as described above. Those identified as having mutations that interact with the subject gene can be tested further to verify the reproducibility and specificity of this genetic interaction.  EP insertions that demonstrate a specific genetic interaction with a mutant or mis-expressed subject gene, have a physically tagged new gene which can be identified and sequenced using PCR or hybridization screening methods, allowing the isolation of the genomic DNA adjacent to the position of the EP element insertion.

EXAMPLES

The following examples describe the isolation and cloning of the nucleic acid sequence of SEQ ID NOS: 1, 3, and 5, and how these sequences, and derivatives and fragments thereof, as well as other pathway nucleic acids and gene products can be used for genetic studies to elucidate mechanisms of a pathway involving a subject protein as well as the discovery of potential pharmaceutical or pesticidal agents that interact with the pathway.

These Examples are provided merely as illustrative of various aspects of the invention and should not be construed to limit the invention in any way.

Example 1: Preparation of Drosophila cDNA Library

A Drosophila expressed sequence tag (EST) cDNA library was prepared as follows. Tissue from mixed stage embryos (0-20 hour), imaginal disks and adult fly heads were collected and total RNA was prepared. Mitochondrial rRNA was removed from the total RNA by hybridization with biotinylated rRNA specific oligonucleotides and the resulting RNA was selected for polyadenylated mRNA. The resulting material was then used to construct a random primed library. First strand cDNA synthesis was primed using a six nucleotide random primer. The first strand cDNA was then tailed with terminal transferase to add approximately 15 dGTP molecules. The second strand was primed using a primer which contained a Not1 site followed by a 13 nucleotide C-tail to hybridize to the G-tailed first strand cDNA. The double stranded cDNA was ligated with BstX1 adaptors and digested with Not1. The cDNA was then fractionated by size by electrophoresis on an agarose gel and the cDNA greater than 700 bp was purified. The cDNA was ligated with Not1, BstX1 digested pCDNA-sk+ vector (a derivative of pBluescript, Stratagene) and used to transform E. coli (XL1blue). The final complexity of the library was 6 X 10^6 independent clones.

The cDNA library was normalized using a modification of the method described by Bonaldo et al. (Genome Research (1996) 6:791-806). Biotinylated driver was prepared from the cDNA by PCR amplification of the inserts and allowed to hybridize with single stranded plasmids of the same library. The resulting double-stranded forms were removed using strepavidin magnetic beads, the remaining
single stranded plasmids were converted to double stranded molecules using Sequenase (Amersham, Arlington Hills, IL), and the plasmid DNA stored at -20°C prior to transformation. Aliquots of the normalized plasmid library were used to transform E. coli (XL1blue or DH10B), plated at moderate density, and the colonies picked into a 384-well master plate containing bacterial growth media using a Qbot robot (Genetix, Christchurch, UK). The clones were allowed to grow for 24 hours at 37°C then the master plates were frozen at -80°C for storage. The total number of colonies picked for sequencing from the normalized library was 240,000. The master plates were used to inoculate media for growth and preparation of DNA for use as template in sequencing reactions. The reactions were primarily carried out with primer that initiated at the 5’ end of the cDNA inserts. However, a minor percentage of the clones were also sequenced from the 3’ end. Clones were selected for 3’ end sequencing based on either further biological interest or the selection of clones that could extend assemblies of contiguous sequences (“contigs”) as discussed below. DNA sequencing was carried out using ABI377 automated sequencers and used either ABI FS, dithodamine or BigDye chemistries (Applied Biosystems, Inc., Foster City, CA).

Analysis of sequences were done as follows: the traces generated by the automated sequencers were base-called using the program “Phred” (Gordon, Genome Res. (1998) 8:195-202), which also assigned quality values to each base. The resulting sequences were trimmed for quality in view of the assigned scores. Vector sequences were also removed. Each sequence was compared to all other fly EST sequences using the BLAST program and a filter to identify regions of near 100% identity.

Sequences with potential overlap were then assembled into contigs using the programs “Phrap”, “Phred” and “Consed” (Phil Green, University of Washington, Seattle, Washington; http://bozeman.mbt.washington.edu/phrap.docs/phrap.html). The resulting assemblies were then compared to existing public databases and homology to known proteins was then used to direct translation of the consensus sequence. Where no BLAST homology was available, the statistically most likely translation based on codon and hexanucleotide preference was used. The Pfam (Bateman et al., Nucleic Acids Res. (1999) 27:260-262) and Prosite (Hofmann et al., Nucleic Acids Res. (1999) 27(1):215-219) collections of protein domains were used to identify motifs in the resulting translations. The contig sequences were archived in an Oracle-based relational database (FlyTag™, Exelixis Pharmaceuticals, Inc., South San Francisco, CA)

**Example 2: Cloning of Subject Nucleic Acid Molecules**

Unless otherwise noted, the PCR conditions used for cloning the dmlKR/SDH nucleic acid sequence was as follows: A denaturation step of 94°C, 5 min; followed by 35 cycles of: 94°C 1 min, 55°C 1 min 72°C 1 min; then, a final extension at 72°C 10 min.
All DNA sequencing reactions were performed using standard protocols for the BigDye sequencing reagents (Applied Biosystems, Inc.) and products were analyzed using ABI 377 DNA sequencers. Trace data obtained from the ABI 377 DNA sequencers was analyzed and assembled into contigs using the Phred-Phrap programs.

Well-separated, single colonies were streaked on a plate and end-sequenced to verify the clones. Single colonies were picked and the enclosed plasmid DNA was purified using Qiagen REAL Preps (Qiagen, Inc., Valencia, CA). Samples were then digested with appropriate enzymes to excise insert from vector and determine size, for example the vector pOT2, (www.fruitfly.org/EST/pOT2vector.html) and can be excised with XhoI/EcoRI; or pBluescript (Stratagene) and can be excised with BssH II. Clones were then sequenced using a combination of primer walking and in vitro transposon tagging strategies.

For primer walking, primers were designed to the known DNA sequences in the clones, using the Primer-3 software (Steve Rozen, Helen J. Skalesky (1998) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/prime3.html). These primers were then used in sequencing reactions to extend the sequence until the full sequence of the insert was determined.

The GPS-1 Genome Priming System in vitro transposon kit (New England Biolabs, Inc., Beverly, MA) was used for transposon-based sequencing, following manufacturer’s protocols. Briefly, multiple DNA templates with randomly interspersed primer-binding sites were generated. These clones were prepared by picking 24 colonies/clone into a Qiagen REAL Prep to purify DNA and sequenced by using supplied primers to perform bidirectional sequencing from both ends of transposon insertion.

Sequences were then assembled using Phred/Phrap and analyzed using Consed. Ambiguities in the sequence were resolved by resequencing several clones. This effort resulted in identification of various nucleic acid molecules, which are described in detail below.

dmLKR/SDH

A dmLKR/SDH nucleic acid molecule was identified in a contiguous nucleotide sequence of 3266 bases in length, encompassing an open reading frame (ORF) of 2784 nucleotides encoding a predicted protein of 928 amino acids. The ORF extends from base 104-2885 of SEQ ID NO:1.

dmPAL

A dmPAL nucleic acid molecule was identified in a contiguous nucleotide sequence of 2652 bases in length, encompassing an open reading frame (ORF) of 1623 nucleotides encoding a predicted protein of 541 amino acids. The ORF extends from base 522-2144 of SEQ ID NO:3.
dmPAL2

A dmPAL2 nucleic acid molecule was identified in a contiguous nucleotide sequence of 1,703 kilobases in length, encompassing an open reading frame (ORF) of 1,218 nucleotides encoding a predicted protein of 406 amino acids. The ORF extends from base 1,133-1,327 of SEQ ID NO:5.

Example 3: Analysis of Subject Nucleic Acid Sequences

Upon completion of cloning, the sequences were analyzed using the Pfam and Prosite programs.

dmLKR/SDH

Pfam predicted an alanine dehydrogenase/pyridine nucleotide transhydrogenase domain at amino acids 23-474 (nucleotides 172-1,525) and a lactate/malate dehydrogenase domain at amino acids 485-498 (nucleotides 1,558-1,597).

Nucleotide and amino acid sequences for the dmLKR/SDH nucleic acid sequences and their encoded proteins were searched against all available nucleotide and amino acid sequences in the public databases, using BLAST (Altschul et al., supra). Table 1 below summarizes the results. The 5 most similar sequences are listed.

<table>
<thead>
<tr>
<th>DNA BLAST</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI#</td>
<td></td>
</tr>
<tr>
<td>6436581= AC014754</td>
<td><em>Drosophila melanogaster</em>, <em><strong>SEQUENCING IN PROGRESS</strong></em>, in ordered pieces</td>
</tr>
<tr>
<td>5670410= AC008328</td>
<td><em>Drosophila melanogaster</em> chromosome 2 clone BACR09A04 (D860) RPCI-98 09.A.4 map 28B-28C strain y; ca bw sp, <em><strong>SEQUENCING IN PROGRESS</strong></em>, 74 unordered pieces</td>
</tr>
<tr>
<td>3492861= AC004722</td>
<td><em>Drosophila melanogaster</em> DNA sequence (P1 DS02760 (D286)), complete sequence</td>
</tr>
<tr>
<td>4445853= AI531718</td>
<td><em>Drosophila melanogaster</em> cDNA clone SD02959 5prime, mRNA sequence</td>
</tr>
<tr>
<td>5670409= AC008329</td>
<td><em>Drosophila melanogaster</em> chromosome 2 clone BACR31D05 (D861) RPCI-98 31.D.5 map 28C-28D strain y; ca bw sp, <em><strong>SEQUENCING IN PROGRESS</strong></em>, 80 unordered pieces</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROTEIN BLAST</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI#</td>
<td></td>
</tr>
<tr>
<td>6094681= AAF03526</td>
<td>Lysine ketoglutarate reductase/saccharopine dehydrogenase [<em>Homo sapiens</em>]</td>
</tr>
<tr>
<td>4107274= CAA12114</td>
<td>Lysine ketoglutarate reductase/saccharopine dehydrogenase [<em>Mus musculus</em>]</td>
</tr>
<tr>
<td>4938304= CAA07619</td>
<td>Lysine-ketoglutarate reductase/saccharopine dehydrogenase [<em>Homo sapiens</em>]</td>
</tr>
<tr>
<td>2736323= AAB94142</td>
<td>Similar to saccharopine dehydrogenases [<em>Caenorhabditis elegans</em>]</td>
</tr>
<tr>
<td>5016095= AAC18622</td>
<td>Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme [<em>Zea mays</em>]</td>
</tr>
</tbody>
</table>
The closest homolog predicted by BLAST analysis is a human LKR/SDH, with 51% identity and 72% homology with dmLKR/SDH. The BLAST analysis also revealed several other LKR/SDH proteins which share significant amino acid homology (40-51% identity; 60-72% similarity) with dmLKR/SDH.

BLAST results for the dmLKR/SDH amino acid sequence indicate 17 amino acid residues as the shortest stretch of contiguous amino acids that is novel with respect to public sequences and 38 amino acids as the shortest stretch of contiguous amino acids for which there are no sequences contained within public database sharing 100% sequence similarity.

dmPAL.

PFAM predicted 2 transmembrane domains at amino acids 16-32 and 464-480 (nucleotides 569-620, and 1913-1964, respectively). PFAM also predicted four NHL consensus domains (PF01436) at amino acids 175-202, 227-255, 283-311, and 387-415 (nucleotides 1046-1130, 1202-1289, 1370-1457, and 1682-1769, respectively).

Nucleotide and amino acid sequences for the dmPAL nucleic acid sequences and their encoded proteins were searched against all available nucleotide and amino acid sequences in the public databases, using BLAST (Altschul et al., supra). Table 2 below summarizes the results. The 5 most similar sequences are listed.

<table>
<thead>
<tr>
<th>DNA BLAST</th>
<th>GI#</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>6437211</td>
<td>AC014124</td>
<td><em>Drosophila melanogaster</em>, <strong>SEQUENCING IN PROGRESS</strong>*, in ordered pieces</td>
</tr>
<tr>
<td>2152680</td>
<td>AA440802</td>
<td><em>Drosophila melanogaster</em> cDNA clone LD15681 5prime, mRNA sequence</td>
</tr>
<tr>
<td>1899683</td>
<td>AA263489</td>
<td><em>Drosophila melanogaster</em> cDNA clone LD06771 5prime, mRNA sequence</td>
</tr>
<tr>
<td>2153463</td>
<td>AA441585</td>
<td><em>Drosophila melanogaster</em> cDNA clone LD16565 5prime, mRNA sequence</td>
</tr>
<tr>
<td>1797119</td>
<td>AA201267</td>
<td><em>Drosophila melanogaster</em> cDNA clone LD04291 5prime, mRNA sequence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROTEIN BLAST</th>
<th>GI#</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>104156</td>
<td>S17855</td>
<td>Peptidylglycine monoxygenase (EC 1.14.17.3) - African clawed frog</td>
</tr>
<tr>
<td>64531</td>
<td>CAA44615</td>
<td>Peptidylhydroxyglycine N-C lyase [Xenopus laevis]</td>
</tr>
<tr>
<td>66354</td>
<td>U13524</td>
<td>Peptidylglycine monoxygenase (EC 1.14.17.3) II precursor - African clawed frog</td>
</tr>
<tr>
<td>214015</td>
<td>AAA49667</td>
<td>Alpha-amidating enzyme precursor [Xenopus laevis]</td>
</tr>
<tr>
<td>5353833</td>
<td>AAD42258</td>
<td>Alpha-amidating enzyme precursor 1 [Lymnaea stagnalis]</td>
</tr>
</tbody>
</table>
The closest homolog predicted by BLAST analysis is a peptidylglycine monooxygenase from African clawed frog, with 45% identity and 64% homology to dmPAL. BLAST searches using dmPAL cDNA and protein sequences show strong homology with the C-terminal PHL domain of a number of bifunctional PAMs from different species.

BLAST results for the dmPAL amino acid sequence indicate 11 amino acid residues as the shortest stretch of contiguous amino acids that is novel with respect to prior art sequences and 15 amino acids as the shortest stretch of contiguous amino acids for which there are no sequences contained within public database sharing 100% sequence similarity.

**dmPAL2**

PFAM predicted four NHL consensus domains (PF01436) at amino acids 179-206, 230-258, 277-305, and 371-403 (nucleotides 647-730, 800-886, 941-1027, and 1223-1321, respectively).

Nucleotide and amino acid sequences for the dmPAL2 nucleic acid sequences and their encoded proteins were searched against all available nucleotide and amino acid sequences in the public databases, using BLAST (Altschul et al., supra). Table 3 below summarizes the results. The 5 most similar sequences are listed.

**TABLE 3**

<table>
<thead>
<tr>
<th>DNA BLAST</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>3928060 = AC005639</td>
<td>Drosophila melanogaster, chromosome 2R, region 59E3-59F4, BAC clone BACR48M01, complete sequence</td>
</tr>
<tr>
<td>6664431 = AC020466</td>
<td>SEQUENCING IN PROGRESS *** Drosophila melanogaster unfinished genome; HTGS phase 2</td>
</tr>
<tr>
<td>2700671 = AA697742</td>
<td>HL03086.5prime HL Drosophila melanogaster head BlueScript Drosophila melanogaster cDNA clone</td>
</tr>
<tr>
<td>326519 = AQ025992</td>
<td>EP(2)0393 Drosophila melanogaster EP line Drosophila melanogaster genomic Sequence recovered from 5' end of P element</td>
</tr>
<tr>
<td>EP465404-A, Claim 9</td>
<td>AE-III (peptidylhydroxyglycine N-C lyase precursor) from Xenopus laevis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROTEIN BLAST</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>163482 = AAA30683</td>
<td>PAM precursor [Bos taurus]</td>
</tr>
<tr>
<td>66352 = URB0AP</td>
<td>peptidylglycine monooxygenase (EC 1.14.17.3) / peptidylamidoglycolate lyase (EC 4.3.2.5) precursor – bovine</td>
</tr>
<tr>
<td>592492 = AAA54099</td>
<td>Sequence 3 from Patent WO 8902460</td>
</tr>
<tr>
<td>66354 = URXLA2</td>
<td>peptidylglycine monooxygenase (EC 1.14.17.3) II precursor - African clawed frog</td>
</tr>
<tr>
<td>214015 = AAA49667</td>
<td>alpha-amidating enzyme precursor [Xenopus laevis] (M19032)</td>
</tr>
</tbody>
</table>
The closest homolog predicted by BLAST analysis is a peptidylglycine monoxygenase from cow with 43% identity and 61% homology to dmPAL2. BLAST searches using dmPAL cDNA and protein sequences show strong homology with the PHL domain of a number of bifunctional PAMs from different species. BLAST results for the dmPAL2 amino acid sequence indicate 12 amino acid residues as the shortest stretch of contiguous amino acids that is novel with respect to published sequences and 18 amino acids as the shortest stretch of contiguous amino acids for which there are no sequences contained within public database sharing 100% sequence similarity.

Example 4: Assay of LKR/SDH activity

Cells expressing recombinant dmLKR/SDH may be lysed. Activity of the LKR/SDH in the supernatant may be measured in a photometric assay in presence or absence of other compounds of interest, using biotinylated NADPH (Sigma, St Louis, MO). Measurements in the decrease of absorbance are taken at 340nm.

Example 5: PAL enzyme activity assay

Cells expressing dmPAL or dmPAL2 are homogenized, and the dmPAL or dmPAL2 protein is extracted with 20mM NaTES, pH 7.4, and 10mM mannitol in presence of protease inhibitors (Husten and Eipper, supra). Insoluble pellets are resuspended in the above mixture with the addition of 1% Triton X-100. The soluble and detergent-extracted proteins are fractionated by gel filtration, and assayed for enzyme activity using the synthetic tripeptide substrates \( \alpha\)-N-acetyl-Tyr-Val_Gly or \( \alpha\)-N-acetyl-Tyr-Val-\( \alpha\)-hydroxyglycine (Perkins et al., Biochem Biophys Res Commun (1990) 171:926-932), in presence or absence of compounds.
CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule of less than about 15 kb in size comprising a nucleic acid sequence that encodes an invertebrate receptor polypeptide and that shares at least about 75% nucleotide sequence identity with the sequence set forth in SEQ ID NO:1, or the complement thereof.

2. An isolated nucleic acid molecule of less than about 15 kb in size comprising a nucleic acid sequence that encodes an invertebrate receptor polypeptide and that shares at least about 75% nucleotide sequence identity with the sequence set forth in SEQ ID NO:3, or the complement thereof.

3. An isolated nucleic acid molecule of less than about 15 kb in size comprising a nucleic acid sequence that encodes an invertebrate receptor polypeptide and that shares at least about 75% nucleotide sequence identity with the sequence set forth in SEQ ID NO:5, or the complement thereof.

4. An isolated nucleic acid molecule comprising a nucleic acid sequence that encodes a polypeptide comprising at least 48 amino acids that share 100% sequence similarity with 48 amino acids of SEQ ID NO:2.

5. The isolated nucleic acid molecule of Claim 4 wherein said nucleic acid sequence encodes the entire sequence of SEQ ID NO:2.

6. An isolated isolated nucleic acid molecule comprising a nucleic acid sequence that encodes the entire sequence of SEQ ID NO:4.

7. The isolated nucleic acid molecule of Claim 6 wherein said nucleic acid sequence encodes at least one NHL domain.

8. An isolated nucleic acid molecule comprising a nucleic acid sequence that encodes a polypeptide comprising at least 28 amino acids that share 100% sequence similarity with 28 amino acids of SEQ ID NO:6.

9. The isolated nucleic acid molecule of Claim 8 wherein said nucleic acid sequence encodes the entire sequence of SEQ ID NO:6.
10. The isolated nucleic acid molecule of Claim 3 wherein said nucleic acid sequence encodes at least NHL domain of SEQ ID NO:6 selected from the group consisting of amino acids residues 179-206, 230-258, 277-305, and 371-403 of SEQ ID NO:6.

11. An isolated nucleic acid molecule of any one of Claims 1-3 wherein said nucleic acid sequence encodes a polypeptide having enzymatic activity.

12. A vector comprising a nucleic acid molecule of any one of Claims 1-11.

13. A host cell comprising a vector of Claim 12.

14. A process for producing an invertebrate enzyme, comprising culturing the host cell of Claim 13 under conditions suitable for expression of said protein and recovering said protein.

15. A purified protein comprising an amino acid sequence having at least about 80% sequence identity with any one of the sequences set forth in SEQ ID NOS:2, 4, and 6.

16. A method for detecting a candidate compound that interacts with an invertebrate enzyme or fragment thereof, said method comprising contacting said enzyme or fragment with one or more candidate molecules; and detecting any interaction between said candidate compound and said enzyme or fragment; wherein the amino acid sequence of said enzyme comprises an amino acid sequence amino acid sequence which is at least about 80% identical to the sequence set forth in any one of SEQ ID NOS:2, 4, and 6.

17. The method of Claim 16 wherein said candidate compound is a putative pesticidal or pharmaceutical agent.

18. The method of Claim 16 wherein said contacting comprises administering said candidate compound to cultured host cells that have been genetically engineered to express said enzyme.

19. The method of Claim 16 wherein said contacting comprises administering said candidate compound to a metazoan invertebrate organism that has been genetically engineered to express said enzyme.
20. The method of Claim 19 wherein said candidate compound is a putative pesticide and said detecting entails observing modulations of a biological activity of said enzyme that result in organism lethality.

5 21. The method of Claim 19 wherein said organism is an insect or worm.

22. A first animal that is an insect or a worm that has been genetically modified to express or mis-express an invertebrate enzyme, or the progeny of said animal that has inherited said enzyme expression or mis-expression, wherein said enzyme comprises an amino acid sequence amino acid sequence which is at least about 80% identical to the sequence set forth in any one of SEQ ID NOS:2, 4, and 6.

23. A method for studying invertebrate enzyme activity comprising detecting the phenotype caused by the expression or mis-expression of said invertebrate enzyme in the first animal of Claim 22.

15 24. A method of controlling a pest, comprising contacting a pest with a compound identified by a method according to claim 16.

25. The method of claim 24, wherein the compound reduces viability of the pest.
<110> GENOPTERA, LLC

<120> Nucleic Acids and Polypeptides of Invertebrate Enzymes and Methods of Use

<130> GNOP-004WO

<150> 60/174,973
<151> 2000-01-06

<150> 60/185,844
<151> 2000-02-29

<150> 60/191,189
<151> 2000-03-22

<160> 6

<170> FastSEQ for Windows Version 4.0

<210> Drosophila melanogaster

<213> DNA

<221> 1

<400> 1

gcagctgggattc ggtatcattcc cactcagggcc aagaataagct acaacggtta caattaaatt 60
cagctgtctct aacatcgcaac taaatcccacc aacagcttggct aacatgtcttg caatgtcttc 120
actggcgcgca aacatcgccgct actgggtgcc acagacacatg ctattgctgggt ctattgccttg 180
tcggcgcgagag gtcaatcctgcg tggggagggcg acagctgtcggc tttggacccaa ccacgggtaca 240
gtgtaataatc gcaacggtttt ctaaggtcatc atggtcagcgc tcacacgctcc gttctctttca 300
cagctgctgcag taccatcggct gctggagttgct tatttctctct gcgtctgctgct 360
gctatatttg ggttaaaacag cgtgacctgatt gctgagttttt gctgacctgatt ctttctcttc 420
cactaatcacc atggccaaag cctctctctct ctgtgttcccc ctgtgttcccc ctgtgttcccc 480
gagatattatt ggtttcattg gatgtggttgc aactttctctct gcgtctgctgct gcgtctgctgct 540
gtcgtttgggg cgtgggtggtt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 600
gtgctttatt gccacggcttt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 660
tcattctcatt tcgtctctct gccgcctggct gcgtctgctgct gcgtctgctgct gcgtctgctgct 720
cattgcgtgct aactttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct gcgtctgctgct 780
cactaatcacc atggccaaag ccatacttatt gcgtctgctgct gcgtctgctgct gcgtctgctgct 840
gagatattatt ggtttcattg gatgtggttgc aactttctctct gcgtctgctgct gcgtctgctgct 900
gtcgtttgggg cgtgggtggtt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 960
gtgctttatt gccacggcttt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1020
cattctcatt tcgtctctct gccgcctggct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1080
gtcgtttgggg cgtgggtggtt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1140
tcattctcatt tcgtctctct gccgcctggct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1200
cactaatcacc atggccaaag ccatacttatt gcgtctgctgct gcgtctgctgct gcgtctgctgct 1260
gagatattatt ggtttcattg gatgtggttgc aactttctctct gcgtctgctgct gcgtctgctgct 1320
gtcgtttgggg cgtgggtggtt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1380
gtgctttatt gccacggcttt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1440
tcattctcatt tcgtctctct gccgcctggct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1500
gtcgtttgggg cgtgggtggtt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1560
tcattctcatt tcgtctctct gccgcctggct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1620
gtgctttatt gccacggcttt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1680
tcattctcatt tcgtctctct gccgcctggct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1740
gtcgtttgggg cgtgggtggtt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1800
gtgctttatt gccacggcttt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1860
tcattctcatt tcgtctctct gccgcctggct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1920
gtcgtttgggg cgtgggtggtt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 1980
gtgctttatt gccacggcttt actttctctct gcgtctgctgct gcgtctgctgct gcgtctgctgct 2040
<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Met</th>
<th>Trp</th>
<th>Arg</th>
<th>Val</th>
<th>Ile</th>
<th>Glu</th>
<th>Leu</th>
<th>Arg</th>
<th>Ala</th>
<th>Thr</th>
<th>Ile</th>
<th>Ala</th>
<th>His</th>
<th>Pro</th>
<th>Phe</th>
<th>Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> 2
<211> 928
<212> PRT
<213> Drosophila melanogaster

<400> 2

Met Trp Arg Val Ile Gln Leu Arg Ala Thr Ile Ala His Pro Phe Thr
1     5   10  15
Arg Gln Arg His Ser Arg Val Ile Ala Arg Glu Asp Gln Ser
20    25  30
Val Trp Glu Arg Ala Pro Phe Gly Pro Thr His Val Gln Lys Leu
35    40  45
Val Lys Gln Asn Val Lys Val Ile Val Gln Pro Ser Asn Arg Arg Ala
50    55  60
Tyr Pro Met Gln Ala Tyr Met Gln Ala Gly Ala His Ile Gln Glu Asp
65    70  75  80
Ile Ser Asp Ala Ser Val Ile Phe Gly Val Lys Gln Val Pro Ile Asp
85    90  95
Ala Leu Ile Pro Gly Lys Thr Tyr Cys Phe Phe Ser His Thr Ile Lys
100   105 110
Ala Gln Glu Ser Asn Met Pro Leu Leu Asp Ala Ile Leu Glu Lys Lys
115   120 125
Ile Arg Leu Ile Asp Tyr Glu Arg Ile Ile Asp Glu Arg Gly Ala Arg
130   135 140
Gln Val Ala Phe Gly Lys Tyr Ala Gly Val Ala Gly Met Val Asn Ile
145   150 155 160
Leu His Gly Ile Gly Leu Arg Leu Ala Leu Gly His His Thr Pro
165   170 175
Phe Met His Ile Gly Pro Ala His Asn Tyr Arg Asn Ser Ser Met Ala
180   185 190
Arg Gln Ala Ile Arg Asp Cys Gly Tyr Glu Ile Ser Leu Gly Met Met
195   200 205
Pro Lys Ser Ile Gly Pro Leu Thr Phe Val Phe Thr Gly Ser Gly Asn
210   215 220
Val Ser Gln Gly Ala Glu Val Phe Ser Glu Leu Pro Ile Gly Tyr
225   230 235 240
Val Pro Pro Glu Met Leu Arg Lys Val Ala Glu Gly Asn Gln Asn
245   250 255
Lys Leu Tyr Gly Cys Glu Val Ser Arg Ser Asp His Leu Glu Arg Arg
260   265 270
Glutamic Acid (Glu), Glycine (Gly), Phenylalanine (Phe), Aspartic Acid (Asp), Alanine (Ala), Lysine (Lys), Tyrosine (Tyr), Asparagine (Asn), Valine (Val), Isoleucine (Ile), Serine (Ser), Threonine (Thr), Proline (Pro), Glutamine (Gln), and Arginine (Arg).
Val Ile His Leu Met Gly Met Ser Asp Ser Thr Ile Phe Tyr Glu Asn
770 775 780
Leu Lys Gln Lys Leu Thr Glu Arg Ile Gly Asp Val Asp Gly Ile Glu
785 790 795 800
Ser Leu Gly Leu Leu Asp Asp Thr Pro Val Val Lys Leu Asn Thr Pro
805 810 815
Leu Asp Thr Leu Ser His Tyr Leu Ser Lys Arg Leu Ala Phe Glu Arg
820 825 830
Asp Glu Arg Asp Leu Val Val Leu Arg His Glu Val Gly Ile Arg Trp
835 840 845
Pro Asp Gly Arg Arg Gly Arg Gly Ile Asn Phe Val Val Tyr Gly
850 855 860
Gln Pro Gln Gly His Ser Ala Met Ala Met Thr Val Gly Lys Pro Ala
865 870 875 880
Ala Ile Ala Ala Lys Met Ile Leu Asp Gly Glu Ile Gln Glu Arg Gly
885 890 895
Val Leu Leu Pro Phe Thr Pro Asp Ile Tyr Arg Pro Met Leu Gln Arg
900 905 910
Leu Arg Ser Glu Gly Leu Thr Ala Thr Glu Thr Ser Arg Trp Leu Asn
915 920 925

<210> 3
<211> 2652
<212> DNA
<213> Drosophila melanogaster

<400> 3
ttcggcacaga ttggctgccgg gttttcttggc ggcgtttttat cattaccgtc agttcctcgtct
60
tgttgacagct aatattttctg gccaacctcttg ccgtcttctat tgtcttaaaaa cacggaagta
120
aacaagatigm aagaacctag aacggcacaac gcacgtacgtat gcacagttgga
gatctcttctgg aatggcttttg gcgtctttcttg ctctggttca
180
cggcggcgca tctgctgtttg ggcgtttttat cattaccgtc agttcctcgtct
240
acgagttttcg tggcgtttttat cattaccgtc agttcctcgtct
300
acgagttttcg tggcgtttttat cattaccgtc agttcctcgtct
360
acgagttttcg tggcgtttttat cattaccgtc agttcctcgtct
420
acgagttttcg tggcgtttttat cattaccgtc agttcctcgtct
480
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
540
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
600
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
660
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
720
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
780
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
840
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
900
tccgggaccct atatatatacg aacggaagct aacggaagct gacggtttttat cattaccgtc agttcctcgtct
960
<210> 4
<211> 541
<212> PRT
<213> Drosophila melanogaster

<400> 4
Met Lys Ser Thr Asp Ser Ala Lys Cys Leu Gly Ser Lys Ser Leu Ala
Ile Cys Leu Leu Leu His Leu Leu Leu Cys Ile Arg Pro Ala Val
Ser Glu Thr Glu Ser Pro Glu Arg Tyr Leu His Asn Val Asp Ser Asn
Ser Asn Asn Asn Glu Arg Leu His Glu Ile Leu Lys Gly Ser Gly Ala
Gly Ser Gly Ala Thr Gin Leu Asn Trp Pro Gin Pro Pro Lys Gin Thr
Val Pro Asn Val Lys Thr Glu Leu Ala Lys Leu Asn Asn Thr Tyr Val
Tyr Gin Asn Ala Trp Pro Ala Asn Asn Val Lys Leu Gly Ala Val Thr
Ala Val Ser Phe Asp Lys Ala Gly Asn Val Val Ile Phe His Arg Val
Asn Arg Val Trp Gly Gin Thr Thr Phe Asp Asn Arg Asn Gin Tyr Gin
Glu Lys Tyr Arg Gly Pro Ile Arg Glu Ser Thr Ile Leu Ala Leu Glu
Pro Ala Thr Gly Lys Val Gin Tyr Asp Trp Gly Lys Asn Phe Phe Tyr
Met Pro His Gly Leu Thr Val Asp Pro Glu Asp Asn Val Trp Leu Thr
Asp Val Ala Met His Gin Val Phe Lys Phe Pro Pro Arg Gly Gly Asp
Gly Lys Pro Ala Leu Thr Leu Gly Asp Ala Phe Gin Pro Gly Ser Gly
210 215 220
Arg Lys Phe Cys Lys Pro Thr Ser Val Ala Val Leu Asp Asn Gly Asp
225 230 235 240
Phe Phe Val Ala Asp Gly Tyr Cys Ala Arg Ile Phe Lys Tyr Ser
245 250 255
Arg Lys Gly Leu Ile Leu Phe Trp Gly Gin Asn Thr Phe Ser Gly
260 265 270
Ile Ser Tyr Val Ala Pro Gin Asn Phe Ala Ile Pro His Ala
275 280 285
Leu Thr Leu Val Pro Glu Leu Gin Leu Leu Cys Ala Ala Asp Arg Glu
290 295 300
Asn Gly Arg Val Gin Cys Phe Leu Ser Ser Asn Gly Thr Phe His Ser
305 310 315 320
Gln Tyr His Asn Gin Leu Ile Gly Asp Arg Leu Phe Ser Met Ala Tyr
325 330 335
Thr Pro Ala Ala Gly Gin Leu Val Ile Val Asn Gin Pro Thr Ala
340 345 350
Glu Leu Gly Ile His Pro Gin Gin His Tyr Asn Gin Val His Gly Phe Val
355 360 365
Leu Ser Met Arg Ser Lys Gln Leu Val Ser Lys Phe Gly Pro Asn Asn
370 375 380
Leu Gln Phe Gln Asn Pro His Asp Val Ala Val Thr Ala Asp Gly Asn
385 390 395 400
Glu Ile Tyr Val Ala Glu Leu Asn Pro Met Arg Ile His Lys Phe Val
405 410 415
His Arg Ser Leu Ala Lys Pro Met Ser Leu Ser Ala Ser Lys Asp Ser
420 425 430
Arg Asp Ser Ala Ile Ser Gln Ala Val Gly Gly Asp Gln Val Pro Ala
435 440 445
Val Ala Val His His Pro Ser Gly Lys Ala Ile Leu Val Ala Ser Leu
450 455 460
Met Leu Phe Ala Gly Ser Thr Phe Ala Leu Ala Leu Ile Phe Ala
465 470 475 480
Arg Arg Arg Lys Gly Cys Leu Pro Phe Gly Ala Arg Gly Arg Arg
485 490 495
His Ala Trp Glu Lys Ser Asp Gly Phe Lys Leu Gly Gly Leu Leu Asp
500 505 510
Arg Asp Arg Asn Gly Phe Glu Leu Asp Gln Gln Ala Asp Ser Glu
515 520 525
Glu Gln Glu Thr Lys Thr Leu Ala Ser Ala Gln Tyr Ala
530 535

<210> 5
<211> 1703
<212> DNA
<213> Drosophila melanogaster

<400> 5
ttcgccccga ggcacgccgg aacagatgtt ggcagttggg aaatacctagt gataagaacc
60
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
120
ttcctctgttc tttgccggttc tagccataga ttgtcgctat gtcctgctca gcagtcgacg
180
cgtcctcttc gcggggttgg ccatgcatgt ggtatcctcttc gcacggctcc
240
tttcctctcag ggcctgctttg ccttcgctcc ggtgcttttg ctgggttctg cttgctttttt
300
ggatgctgag cagctctctcc cagctctctc cagctctctc cagctctctc cagctctctc
360
agcagctgct cacatcctcttc gcctcagctt ggtatcctcttc gcacggctcc
420
gagcagttgg cgcctgctttg ccttcgctcc ggtgcttttg ctgggttctg cttgctttttt
480
ggatgctgag cagctctctcc cagctctctc cagctctctc cagctctctc cagctctctc
540
ggatgctgag cagctctctcc cagctctctc cagctctctc cagctctctc cagctctctc
600
ttcgccccga ggcacgccgg aacagatgtt ggcagttggg aaatacctagt gataagaacc
660
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
720
ttcctctgttc tttgccggttc tagccataga ttgtcgctat gtcctgctca gcagtcgacg
780
agcagctgct cacatcctcttc gcctcagctt ggtatcctcttc gcacggctcc
840
gagcagttgg cgcctgctttg ccttcgctcc ggtgcttttg ctgggttctg cttgctttttt
900
ggatgctgag cagctctctcc cagctctctc cagctctctc cagctctctc cagctctctc
960
ttcgccccga ggcacgccgg aacagatgtt ggcagttggg aaatacctagt gataagaacc
1020
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
1080
ttcctctgttc tttgccggttc tagccataga ttgtcgctat gtcctgctca gcagtcgacg
1140
ttcctttcag cagactccag ctaggtgtcc gtagctttaa ttaagttttac gaaaccctgg
1200
ttcctttcag cagactccag ctaggtgtcc gtagctttaa ttaagttttac gaaaccctgg
1260
ttcctttcag cagactccag ctaggtgtcc gtagctttaa ttaagttttac gaaaccctgg
1320
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
1380
ttcctctgttc tttgccggttc tagccataga ttgtcgctat gtcctgctca gcagtcgacg
1440
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
1500
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
1560
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
1620
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
1680
agaacccgagt cgcctgttgg ttggttgaag aatacctagc accctcagaga aatttgcgcg
1703

<210> 6
<211> 406
<212> FRT
<213> Drosophila melanogaster
<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Met Ser Arg Leu Leu Phe Val Ala Leu Ala Leu Ala Ile Ser Leu Gly Tyr</td>
</tr>
<tr>
<td>2</td>
<td>Val Ala Ser Ser Ser Ser Asn His Leu Pro Ala Gly Leu Ala Met Asp</td>
</tr>
<tr>
<td>3</td>
<td>Leu Gly Pro Gly Val Asn Leu Asn Glu Arg Phe Phe Asp Gln Val Arg</td>
</tr>
<tr>
<td>4</td>
<td>Ala Leu Ile Lys Arg Arg Leu Gln Glu Lys Gly Leu Ala Lys Pro Glu</td>
</tr>
<tr>
<td>5</td>
<td>Gln Pro Glu Leu Ala Met Pro Leu Thr Asp Asp Asp Ala Val Ala Leu</td>
</tr>
<tr>
<td>6</td>
<td>Gln Asn Gln Arg Ser Tyr Asp Asn Val Pro Leu Pro Ala Ala Ser Val</td>
</tr>
<tr>
<td>7</td>
<td>Pro Thr Pro Val Leu Val Glu Asn Trp Pro Thr Glu Gln His Ser Phe</td>
</tr>
<tr>
<td>8</td>
<td>Gly Gln Val Thr Ala Val Ala Val Asp Pro Gln Gly Ser Pro Val Val</td>
</tr>
<tr>
<td>9</td>
<td>Phe His Arg Ala Glu Arg Tyr Trp Asp Val Asn Thr Phe Asn Glu Ser</td>
</tr>
<tr>
<td>10</td>
<td>Asn Ile Tyr Tyr Leu Ile Glu Tyr Gly Pro Ile Lys Glu Asn Thr Ile</td>
</tr>
<tr>
<td>11</td>
<td>Tyr Val Leu Asp Ala Lys Thr Gly Ala Ile Lys Ser Gly Trp Gly Ser</td>
</tr>
<tr>
<td>12</td>
<td>Asn Met Phe Tyr Met Pro His Gly Leu Thr Ile Asp Leu His Gly Asn</td>
</tr>
<tr>
<td>13</td>
<td>Tyr Trp Ile Thr Asp Val Ala Met His Gln Ala Phe Lys Phe Lys Pro</td>
</tr>
<tr>
<td>14</td>
<td>Phe Ser Asn Lys Pro Leu Leu Thr Ile Gly Lys Arg Phe Arg Pro Gly</td>
</tr>
<tr>
<td>15</td>
<td>Ser Ser Val Lys His Leu Cys Lys Pro Thr Ser Ile Ala Val Ala Thr</td>
</tr>
<tr>
<td>16</td>
<td>Thr Gly Glu Phe Phe Ile Ala Asp Gly Tyr Cys Asn Ser Arg Ile Leu</td>
</tr>
<tr>
<td>17</td>
<td>Lys Phe Asn Ala Ala Gly Lys Leu Leu Arg Thr Ile Pro Gln Pro Pro</td>
</tr>
<tr>
<td>18</td>
<td>Glu Phe Leu Ser Leu Gln Val Pro His Ala Ile Thr Leu Leu Glu His</td>
</tr>
<tr>
<td>19</td>
<td>Leu Asp Leu Leu Cys Ile Ala Asp Arg Glu Asn Met Arg Val Val Cys</td>
</tr>
<tr>
<td>20</td>
<td>Pro Lys Ala Gly Leu Ile Ser Ser His Gly Glu Gly Glu Pro Ala Ala</td>
</tr>
<tr>
<td>21</td>
<td>Thr Ile Gln Glu Pro Asp Leu Gly Arg Val Phe Gly Val Ala Ser Phe</td>
</tr>
<tr>
<td>22</td>
<td>Gly Asp Ile Val Phe Ala Val Asn Gly Pro Thr Ser Met Leu Pro Val</td>
</tr>
<tr>
<td>23</td>
<td>Arg Gly Phe Thr Ile Asp Pro Arg Ser Glu Thr Ile Ile Gly His Trp</td>
</tr>
<tr>
<td>24</td>
<td>Gly Glu Phe Lys Asn Pro His Ser Met Ala Val Ser Val Asn Gly Ser</td>
</tr>
<tr>
<td>25</td>
<td>Ala Leu Tyr Val Thr Glu Ile Gly Thr Asn His Glu Thr Asn Arg Val</td>
</tr>
<tr>
<td>26</td>
<td>Trp Lys Tyr Val Leu Ala</td>
</tr>
</tbody>
</table>

-7-