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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2003 (30.05.2003)

PCT

(10) International Publication Number
WO 03/044212 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/US02/36599

(22) International Filing Date: 14 November 2002 (14.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/332,947 16 November 2001 (16.11.2001) US

(71) Applicant (*for all designated States except US*): **EXELIXIS, INC.** [US/US]; P.O. Box 511, 170 Harbor Way, South San Francisco, CA 94083-0511 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **ROBERTSON, Stephanie, A.** [US/US]; 255 Fowler Avenue, San Francisco, CA 94127 (US). **OLLMANN, Michael, Martin** [US/US]; 8 Odessa Court, Redwood City, CA 94063 (US). **SONG, Chunyan** [CN/US]; 95 Behr Avenue #103, San Francisco, CA 94131 (US). **KEEGAN, Kevin, Patrick** [US/US]; 8989 Renato Street, San Diego, CA 92129 (US).

(74) Agents: **SHAYESTEH, Laleh** et al.; Exelixis, Inc., P.O. Box 511, 170 Harbor Way, South San Francisco, CA 94083-0511 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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.

WO 03/044212 A2

(54) **Title:** NUCLEIC ACIDS AND POLYPEPTIDES OF INVERTEBRATE BRCA2 AND METHODS OF USE

(57) **Abstract:** BRCA2 (BRCA2) nucleic acids and proteins that have been isolated from *Drosophila melanogaster* and *Tribolium castaneum* are described. The BRCA2 nucleic acids and proteins can be used to genetically modify metazoan invertebrate organisms or cultured cells, resulting in BRCA2 expression or mis-expression. 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 BRCA2 protein. They can also be used in methods for studying BRCA2 activity and identifying other genes that modulate the function of, or interact with, the BRCA2 gene.

NUCLEIC ACIDS AND POLYPEPTIDES OF INVERTEBRATE BRCA2 AND METHODS OF USE

REFERENCE TO RELATED APPLICATION

This application claims priority to US provisional patent application 60/332,947, filed 11/16/2001. The contents of the prior application are hereby incorporated in its 5 entirety.

BACKGROUND OF THE INVENTION

BRCA2 is a tumor suppressor gene that was identified (Wooster R, et al., (1995) Nature 378:789-792; Tavtigian S, et al., (1996) Nat Genet 12:333-337) in families 10 predisposed to breast cancer. In addition to familial breast cancer, BRCA2 has also been linked to other cancers including canceros of ovary, hepatocellular, pancreas, and prostate (Gudmundsson J, et al., (1996) Am J Hum Genet 58:749-756; Thorlacius S, et al., (1996) Nat Genet 13:117-119; Katagiri T., et al., (1996) Cancer Res 56:4575-4577; Goggins M, et al., (1996) Cancer Res 56:5360-5364). Human BRCA2 gene encodes a large protein 15 of 3418 amino acids (Tavtigian, *supra*), is highly expressed in rapidly proliferating cells, with the expression peaking at the G1/S boundary of the cell cycle. BRCA2-null mouse embryos are nonviable at a very early stage of development and blastocysts derived from these embryos are very sensitive to gamma-irradiation (Sharan, S. K., et al. (1997) Nature 386: 804-810). Furthermore, BRCA2 interacts with the DNA repair protein Rad51 20 (Sharan, *supra*), specifically at the BRC repeats (Chen P-L, et al., (1998) PNAS 95:5287-5292), and expression of BRC repeats in breast cancer cells disrupts the BRCA2-Rad51 complex and leads to radiation hypersensitivity and loss of G2/M checkpoint control (Chen C-F; et al. (1999) J Biol Chem 274: 32931-32935). Thus, BRCA2 plays a crucial role in cellular response to DNA damage.

25 Although several mammalian and vertebrate homologs of the tumor suppressor BRCA2 have been described, only one invertebrate homolog has been identified to date in the worm *C.elegans* (T07E3.5; Koonin EV, et al., (1996) Nat Gent 13:266-268). Few lines of evidence, however, have hinted at the existence of a BRCA2 homolog in any other invertebrate species, such as the fruit fly *Drosophila*. Indeed, a likely ortholog of BRCA2 30 was not found after a detailed analysis of the sequenced genome of *Drosophila* (Rubin GM, et al., (2000) Science 287:2204-2215).

Identification of novel BRCA2 orthologues in model organisms such as *Drosophila melanogaster* and other insect species provides important and useful tools for genetic and molecular study and validation of these molecules as potential pharmaceutical targets.

5 The present invention discloses novel compound targets which are members of the BRCA2 class of proteins from the fruit fly *Drosophila melanogaster* and the red flour beetle *tribolium castaneum.*

SUMMARY OF THE INVENTION

10 It is an object of the present invention to provide invertebrate homologs of a BRCA2, that can be used in genetic screening methods to characterize pathways that BRCA2 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 BRCA2 such as those that may have utility as therapeutics.

15 These and other objects are provided by the present invention which concerns the identification and characterization of novel BRCA2 in *Drosophila melanogaster* and *Tribolium castaneum.* Isolated nucleic acid molecules are provided that comprise nucleic acid sequences encoding BRCA2 protein as well as novel fragments and derivatives thereof. Vectors and host cells comprising the BRCA2 nucleic acid molecules are also

20 described, as well as metazoan invertebrate organisms (*e.g.* insects, coelomates and pseudocoelomates) that are genetically modified to express or mis-express a BRCA2 protein.

An important utility of the novel BRCA2 nucleic acids and proteins is that they can be used in screening assays to identify candidate compounds which are potential

25 therapeutics that interact with BRCA2 proteins. Such assays typically comprise contacting a BRCA2 protein or fragment with one or more candidate molecules, and detecting any interaction between the candidate compound and the BRCA2 protein. The assays may comprise adding the candidate molecules to cultures of cells genetically engineered to express BRCA2 proteins, or alternatively, administering the candidate

30 compound to a metazoan invertebrate organism genetically engineered to express BRCA2 protein.

The genetically engineered metazoan invertebrate animals of the invention can also be used in methods for studying BRCA2 activity. These methods typically involve detecting the phenotype caused by the expression or mis-expression of the BRCA2

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 BRCA2 protein.

5

DETAILED DESCRIPTION OF THE INVENTION

The use of invertebrate model organism genetics and related technologies can greatly facilitate the elucidation of biological pathways (Scangos, Nat. Biotechnol. (1997) 15:1220-1221; Margolis and Duyk, *supra*). Of particular use is the insect model
10 organism, *Drosophila melanogaster* (hereinafter referred to generally as "*Drosophila*"). An extensive search for BRCA2 nucleic acids and their encoded proteins in *Drosophila* and the red flour beetle *Tribolium castaneum* (hereinafter referred to as *Tribolium*) was conducted in an attempt to identify new and useful tools for probing the function and regulation of the BRCA2 genes, and for use as targets in drug discovery.

15 Novel BRCA2 nucleic acids and their encoded protein are identified herein. The newly identified BRCA2 nucleic acids can be used for the generation of mutant phenotypes in animal models or in living cells that can be used to study regulation of BRCA2, and the use of BRCA2 as a drug target. Due to the ability to rapidly carry out large-scale, systematic genetic screens, the use of invertebrate model organisms such as
20 *Drosophila* has great utility for analyzing the expression and mis-expression of a BRCA2 protein. Thus, the invention provides a superior approach for identifying other components involved in the synthesis, activity, and regulation of BRCA2 proteins. Systematic genetic analysis of BRCA2s using invertebrate model organisms can lead to the identification and validation of compound targets directed to components of the
25 BRCA2 pathway. Model organisms or cultured cells that have been genetically engineered to express BRCA2 can be used to screen candidate compounds for their ability to modulate BRCA2 expression or activity, and thus are useful in the identification of new drug targets, therapeutic agents, diagnostics and prognostics useful in the treatment of disorders of abnormal cell cycle checkpoint activity, such as cnacer. The details of the
30 conditions used for the identification and/or isolation of novel BRCA2 nucleic acids and proteins are described in the Examples section below. Various non-limiting embodiments of the invention, applications and uses of these novel BRCA2 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.

BRCA2 Nucleic Acids

The invention relates generally to nucleic acid sequences of BRCA2s, and more particularly BRCA2 nucleic acid sequences of *Drosophila* and *Tribolium*, and methods of using these sequences. As described in the Examples below, nucleic acid sequences were isolated from *Drosophila* (SEQ ID NOs:1, 3) and *Tribolium* (SEQ ID NO:5) that encode BRCA2 homologs, hereinafter referred to as DBRCA2.1 for SEQ ID NO:1, DBRCA2.2 for SEQ ID NO:3, and TBRCA2 for SEQ ID NO:5. 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 BRCA2 nucleic acid sequences 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. BRCA2 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 SEQ ID NOs:1, 3, and 5 have utility in inhibiting the function of BRCA2 proteins. The fragments are of length sufficient to specifically hybridize with the corresponding 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 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 5kb, more preferably less than 2 kb, and in some cases, preferably less than 500 bases.

Additional preferred fragments of SEQ ID NO:1 encode the BRC repeats, which are located at approximately nucleotides 1772-1877, 2075-2180, and 2300-2405, respectively.

Additional preferred fragments of SEQ ID NO:3 encode the BRC repeats, which are located at approximately nucleotides 1635-1740, 1938-2043, and 2163-2268, respectively.

Additional preferred fragments of SEQ ID NO:5 encode the BRC repeats, which
5 are located at approximately nucleotides 369-468, 472-576, and 592-696, respectively.

The subject nucleic acid sequences may consist solely of SEQ ID NOs:1, 3, or 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.
10 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
15 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 sequences of BRCA2 include sequences that hybridize to the nucleic acid sequence of SEQ ID NOs:1, 3, or 5 under stringency conditions such that
20 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
25 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
30 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 SEQ ID NOS:1, 3, or 5 under highly stringent hybridization conditions that are: prehybridization of filters containing nucleic acid for 8 hours to 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 1h in a solution containing 0.1X SSC and 0.1% SDS (sodium dodecyl sulfate).

Derivative nucleic acid sequences that have at least about 70% sequence identity with SEQ ID NOs:1, 3, or 5are capable of hybridizing to SEQ ID NOs:1, 3, or 5under moderately stringent conditions that are: 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 SEQ ID NOs:1, 3, or 5under 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.

Derivative BRCA2 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 NOS:1, 3, or 5, or domain-encoding regions thereof.

In one preferred embodiment, the derivative nucleic acid encodes a polypeptide comprising a BRCA2 amino acid sequence of any of SEQ ID NOs:2, 4, or 6 or a fragment or derivative thereof as described further below under the subheading "BRCA2 proteins". A derivative BRCA2 nucleic acid sequence, or fragment thereof, may comprise 100% sequence identity with SEQ ID NOS:1, 3, or 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 (Wada *et al.*, Nucleic Acids Research (1990) 18(Suppl.):2367-2411). A detailed discussion of the humanization of nucleic acid sequences is provided in U.S. Pat. No. 5,874,304 to Zolotukhin *et al.* 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 BRCA2 proteins by using specific codons chosen according to the preferred codons used in highly expressed genes in each organism.

More specific embodiments of preferred BRCA2 protein fragments and derivatives are discussed further below in connection with specific BRCA2 proteins.

Isolation, Production, And Expression of BRCA2 Nucleic Acids

Nucleic acid encoding the amino acid sequence of SEQ ID NOs:2, 4, or 6, or fragment or derivative thereof, may be obtained from an appropriate cDNA library prepared from any eukaryotic species that encodes BRCA2 proteins 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*, *nematode*s, 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 an isolated 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 nucleic acids of the BRCA2 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 BRCA2 nucleic acids and derivatives 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 antisense BRCA2 nucleic acid sequences. Newly generated sequences may be identified and isolated using standard methods.

An isolated BRCA2 nucleic acid sequence 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 BRCA2 nucleic acid. Suitable methods for isolating and producing the subject nucleic acid sequences are well-known in the art (Sambrook *et al.*, *supra*; DNA Cloning: A Practical Approach, Vol. 1, 2, 3, 4, (1995) Glover, ed., MRL Press, Ltd., Oxford, U.K.).

5 The nucleotide sequence encoding a BRCA2 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 BRCA2 gene and/or its flanking regions. A variety of host-vector systems may be utilized to express the 10 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 BRCA2 protein may be controlled by a suitable promoter/enhancer element. In addition, a host cell strain 15 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 the BRCA2 gene product, the expression vector can comprise a promoter operably linked to a BRCA2 gene nucleic acid, one or more origins of replication, and, one or more selectable markers (*e.g.* thymidine kinase activity, 20 resistance to antibiotics, *etc.*). Alternatively, recombinant expression vectors can be identified by assaying for the expression of the BRCA2 gene product based on the physical or functional properties of the BRCA2 protein in *in vitro* assay systems (*e.g.* immunoassays).

The BRCA2 protein, fragment, or derivative may be optionally expressed as a 25 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 30 use of a peptide synthesizer.

Once a recombinant that expresses the BRCA2 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 BRCA2 proteins can be purified from natural sources, by standard methods (*e.g.* immunoaffinity purification).

5

BRCA2 Proteins

BRCA2 proteins of the invention comprise or consist of an amino acid sequence of SEQ ID NOs:2, 4, or 6, or fragments or derivatives thereof. Compositions comprising these proteins may consist essentially of the BRCA2 protein, fragments, or derivatives, or
10 may comprise additional components (*e.g.* pharmaceutically acceptable carriers or excipients, culture media, *etc.*).

BRCA2 protein derivatives typically share a certain degree of sequence identity or sequence similarity with 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
15 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
20 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
25 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
30 polar amino acids are glutamine and asparagine; interchangeable basic amino acids arginine, lysine and histidine; interchangeable acidic amino acids aspartic acid and glutamic acid; and interchangeable small amino acids alanine, serine, cystine, threonine, and glycine.

In one preferred embodiment, a BRCA2 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 5 amino acids, and in some cases, the entire length of any of SEQ ID NOs:2, 4, or 6.

The fragment or derivative of the BRCA2 protein is preferably "functionally active" meaning that the BRCA2 protein derivative or fragment exhibits one or more functional activities associated with a full-length, wild-type BRCA2 protein comprising the amino acid sequence of SEQ ID NOs:2, 4, or 6. As one example, a fragment or 10 derivative may have antigenicity such that it can be used in immunoassays, for immunization, for inhibition of BRCA2 activity, *etc*, as discussed further below regarding generation of antibodies to BRCA2 proteins. Preferably, a functionally active BRCA2 fragment or derivative is one that displays one or more biological activities associated with BRCA2 proteins, such as interaction with Rad51. For purposes herein, functionally active 15 fragments also include those fragments that exhibit one or more structural features of a BRCA2, such as BRC repeats, which are located at amino acids 239-273, 340-374, and 415-449 for SEQ ID NO:2, amino acids 546-580, 647-681, and 722-756 for SEQ ID NO:4, and amino acids 124-156, 158-192, and 198-232 for SEQ ID NO:6. The functional activity of BRCA2 proteins, derivatives and fragments can be assayed by various methods 20 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 BRCA2 protein).

BRCA2 derivatives can be produced by various methods known in the art. The 25 manipulations which result in their production can occur at the gene or protein level. For example, a cloned BRCA2 gene sequence can be cleaved at appropriate sites with restriction endonuclease(s) (Wells *et al.*, Philos. Trans. R. Soc. London SerA (1986) 317:415), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*, and expressed to produce the desired derivative. Alternatively, a BRCA2 gene can 30 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 BRCA2 protein sequence.

Chimeric or fusion proteins can be made comprising a BRCA2 protein or fragment thereof (preferably comprising one or more structural or functional domains of the BRCA2 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 a BRCA2-coding sequence 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 chimeric product; and protein synthetic techniques, *e.g.* by use of a peptide synthesizer.

BRCA2 Gene Regulatory Elements

BRCA2 gene regulatory DNA elements, such as enhancers or promoters, can be used to identify tissues, cells, genes and factors that specifically control BRCA2 protein production. Analyzing components that are specific to BRCA2 protein function can lead to an understanding of how to manipulate these regulatory processes, especially therapeutic applications, as well as an understanding of how to diagnose dysfunction in these processes.

Gene fusions with the BRCA2 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 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 BRCA2 proteins (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 BRCA2 proteins 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 BRCA2 proteins.

Alternatively, a binary reporter system can be used, similar to that described further below, where the BRCA2 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 BRCA2 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 UAS_G 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.

BRCA2 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 BRCA2 genes, or promoting the growth and differentiation of the tissues that expresses the BRCA2 protein. BRCA2 gene regulatory DNA elements are also useful in protein-DNA binding assays to identify gene regulatory

proteins that control the expression of BRCA2 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 BRCA2 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 BRCA2 gene regulatory DNA elements. For example, a *Drosophila* cDNA library in an expression vector, can be screened for cDNAs that encode BRCA2 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).

Identification of Molecules that Interact With BRCA2

A variety of methods can be used to identify or screen for molecules, such as proteins or other molecules, that interact with BRCA2 protein, or derivatives or fragments thereof. The assays may employ purified BRCA2 protein, or cell lines or model organisms such as *Drosophila* and *C. elegans*, that have been genetically engineered to express BRCA2 protein. Suitable screening methodologies are well known in the art to test for proteins and other molecules that interact with BRCA2 gene and protein (see *e.g.*, PCT International Publication No. WO 96/34099). The newly identified interacting molecules may provide new targets for pharmaceutical 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, the BRCA2 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

nondenaturing polyacrylamide gel electrophoresis), Western analysis, non-denaturing gel electrophoresis, *etc.*

Another preferred method for identifying interacting proteins is a two-hybrid assay system or variation thereof (Fields and Song, Nature (1989) 340:245-246; U.S. Pat. No. 5 5,283,173; for review see Brent and Finley, Annu. Rev. Genet. (1997) 31:663-704).

Antibodies and immunoassays

BRCA2 proteins encoded by SEQ ID NOs:2, 4, or 6, and derivatives and fragments thereof, such as those discussed above, may be used as an immunogen to 10 generate monoclonal or polyclonal antibodies and antibody fragments or derivatives (*e.g.* chimeric, single chain, Fab fragments). For example, fragments of a BRCA2 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 15 *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, BRCA2 polypeptide fragments provide specific antigens and/or immunogens, especially when coupled to carrier proteins. For example, peptides are covalently coupled to keyhole 20 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* 25 *vitro* binding assays, *etc.* Binding affinity may be assayed by determination of equilibrium constants of antigen-antibody association (usually at least about 10^7 M^{-1}, preferably at least about 10^8 M^{-1}, more preferably at least about 10^9 M^{-1}).

Immunoassays can be used to identify proteins that interact with or bind to BRCA2 protein. Various assays are available for testing the ability of a protein to bind to or 30 compete with binding to a wild-type BRCA2 protein or for binding to an anti-BRCA2 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 Drug Targets

5 Once new BRCA2 genes or BRCA2 interacting genes are identified, they can be assessed as potential drug targets.

Putative 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) BRCA2 activity can be observed. Alternatively, the effect of various 10 compounds on BRCA2s can be assayed using cells that have been engineered to express one or more BRCA2s and associated proteins.

Assays of Compounds on Insects

Potential compounds can be administered to insects in a variety of ways, including 15 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 20 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 25 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 30 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, behavior and neurological defects are observed, and effects on 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, neurological defects, and death.

Assay of Compounds using Cell Cultures

Compounds that modulate (*e.g.* block or enhance) BRCA2 activity may also be assayed using cell culture. For example, various compounds added to cells expressing BRCA2 may be screened for their ability to modulate the activity of BRCA2 genes based upon measurements of cell survival or activity in cell cycle. Assays can be performed on cultured cells expressing endogenous normal or mutant BRCA2s. Such studies also can be performed on cells transfected with vectors capable of expressing the BRCA2s, or functional domains of one of the BRCA2s, in normal or mutant form. In addition, to enhance the signal measured in such assays, cells may be cotransfected with genes encoding BRCA2 proteins.

For example, cell clones are generated that have tetracycline-inducible expression of BRC repeats, with GFP as markers of BRC activity (Chen C-F, et al., (1999) Biol Chem 274:32931-32935). Expression of wild type or mutant BRC is induced, in absence or presence of compounds, followed by gamma-irradiation of cells, incubation for 2 weeks. Cells are then fixed and counted.

Compounds that selectively modulate the BRCA2 are identified as potential drug candidates having BRCA2 specificity.

Identification of small molecules and compounds as potential 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 BRCA2 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.*

Generation and Genetic Analysis of Animals and Cell Lines with Altered Expression of BRCA2 Gene

Both genetically modified animal models (i.e. *in vivo* models), such as *Drosophila*, and *in vitro* models such as genetically engineered cell lines expressing or mis-expressing BRCA2 pathway genes, are useful for the functional analysis of these proteins. Model systems that display detectable phenotypes, can be used for the identification and
5 characterization of BRCA2 pathway genes or other genes of interest and/or phenotypes associated with the mutation or mis-expression of BRCA2 pathway protein. The term "mis-expression" as used herein encompasses mis-expression due to gene mutations. Thus, a mis-expressed BRCA2 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-
10 expressed BRCA2 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
15 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 BRCA2 pathway gene, or any other gene involved in regulation or modulation, or downstream effector of the BRCA2 pathway.

The *in vivo* and *in vitro* models may be genetically engineered or modified so that
20 they 1) have deletions and/or insertions of one or more BRCA2 pathway genes, 2) harbor interfering RNA sequences derived from BRCA2 pathway genes, 3) have had one or more endogenous BRCA2 pathway genes mutated (*e.g.* contain deletions, insertions, rearrangements, or point mutations in BRCA2 gene or other genes in the pathway), and/or 4) contain transgenes for mis-expression of wild-type or mutant forms of such genes.
25 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 BRCA2 pathway gene and/or gene products, and also downstream effectors of BRCA2 function, genes regulated by BRCA2, etc. The model systems can be used for testing potential pharmaceutical compounds that interact with the BRCA2 pathway, for example by
30 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

5 Loss-of-function mutations in an invertebrate metazoan BRCA2 gene can be generated by any of several mutagenesis methods known in the art (Ashburner, In Drosophila melanogaster: A Laboratory Manual (1989) , Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: pp. 299-418; Fly pushing: The Theory and Practice of Drosophila melanogaster Genetics (1997) Cold Spring Harbor Press, Plainview, NY; The 10 nematode *C. elegans* (1988) Wood, Ed., Cold Spring Harbor Laboratory Press, Cold Spring harbor, New York). Techniques for producing mutations in a gene or genome include 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 15 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; gene targeting by homologous recombination; antisense; double-stranded RNA interference; peptide and RNA aptamers; directed deletions; homologous 20 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 25 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 30 (1991) 88:2731-2735) or by inverse polymerase chain reaction (Rehm, website at www.fruitfly.org/methods

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).

Gene targeting approaches using homologous recombination have proven to be successful in *Drosophila* (Rong and Golic, Science (2000) 288:2013-20018) and potentially provide a general method of generating directed mutations in any gene-of-interest. This method uses broken-ended extrachromosomal DNA, created in vivo, to produce homology-directed changes in a target locus. First, a "targeting construct" is designed for the gene-of-interest which allows the replacement of the normal endogenous gene with a specifically designed mutation, such as a deletion, insertion or point mutation, via homologous recombination. The targeting construct is typically carried in an appropriate transposon-mediated transgenesis vector (e.g. P element-, piggyBac-, hermes-, minos-, or mariner-based vectors) which inserts the targeting construct randomly within the genome of the organism. The targeting construct is converted to a recombinogenic extrachromosomal form by inducing the expression of separate transgenes encoding a site-specific recombinase (e.g. FLP, cre, Kw, etc.) which excises the targeting construct, and a rare-cutting site-specific endonuclease (e.g. SceI, CreI, HO, etc.) which generates recombinogenic ends that direct homologous recombination and gene replacement of the endogenous locus. Though this method has only been shown to work in Dros, it has application to worms, other animals, plants, algae etc.

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

BRCA2 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 BRCA2 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 including cactus, pecanex, and Krüppel (LaBonne *et al.*, Dev. Biol. (1989) 136(1):1-16; Schuh and Jackle, Genome (1989) 31(1):422-425; Geisler *et al.*, Cell (1992) 71(4):613-621).

Loss-of-function phenotypes can also be generated by cosuppression methods (Bingham Cell (1997) 90(3):385-387; Smyth, Curr. Biol. (1997) 7(12):793-795; Que and

Jorgensen, Dev. Genet. (1998) 22(1):100-109). Cosuppression is a phenomenon of reduced gene expression produced by expression or injection of a sense strand RNA corresponding to a partial segment of the gene of interest. Cosuppression effects have been employed extensively in plants and *C. elegans* to generate loss-of-function phenotypes, and there is a single report of cosuppression in *Drosophila*, where reduced expression of the Adh gene was induced from a white-Adh transgene using cosuppression methods (Pal-Bhadra *et al.*, Cell (1997) 90(3):479-490).

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 BRCA2 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 containg a sufficient concentration of the dsRNA. In another embodiment of the method, dsRNA derived from BRCA2 genes can be generated *in vivo* by simultaneous expression of both sense and antisense RNA from appropriately positioned promoters operably fused to BRCA2 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 *Heliothis*.

Recently, RNAi has been successfully used in cultured *Drosophila* cells to inhibit expression of targeted proteins (Dixon lab, University of Michigan, http://dixonlab.biochem.med.umich.edu/protocols/RNAiExperiments.html). Thus, cell lines in culture can be manipulated using RNAi both to perturb and study the function of

BRCA2 pathway components and to validate the efficacy of therapeutic strategies which involve the manipulation of this pathway.

Generating Loss-of-function Phenotypes Using Peptide and RNA Aptamers

5 Another method for generating loss-of-function phenotypes is by the use of peptide aptamers, which are peptides or small polypeptides that act as dominant inhibitors of protein function. Peptide aptamers specifically bind to target proteins, blocking their function ability (Kolonin and Finley, PNAS (1998) 95:14266-14271). Due to the highly selective nature of peptide aptamers, they may be used not only to target a specific protein, 10 but also to target specific functions of a given protein (*e.g.* involvement in cellular response to DNA damage). Further, peptide aptamers may be expressed in a controlled fashion by use of promoters which regulate expression in a temporal, spatial or inducible manner. Peptide aptamers act dominantly; therefore, they can be used to analyze proteins for which loss-of-function mutants are not available.

15 Peptide aptamers that bind with high affinity and specificity to a target protein may be isolated by a variety of techniques known in the art. In one method, they are isolated from random peptide libraries by yeast two-hybrid screens (Xu *et al.*, PNAS (1997) 94:12473-12478). They can also be isolated from phage libraries (Hoogenboom *et al.*, Immunotechnology (1998) 4:1-20) or chemically generated peptides/libraries.

20 RNA aptamers are specific RNA ligands for proteins, that can specifically inhibit protein function of the gene (Good *et al.*, Gene Therapy (1997) 4:45-54; Ellington. *et al.*, Biotechnol. Annu. Rev. (1995) 1:185-214). *In vitro* selection methods can be used to identify RNA aptamers having a selected specificity (Bell *et al.*, J. Biol. Chem. (1998) 273:14309-14314). It has been demonstrated that RNA aptamers can inhibit protein 25 function in *Drosophila* (Shi *et al.*, Proc. Natl. Acad. Sci USA (19999) 96:10033-10038). Accordingly, RNA aptamers can be used to decrease the expression of BRCA2 protein or derivative thereof, or a protein that interacts with the BRCA2 protein.

Transgenic animals can be generated to test peptide or RNA aptamers *in vivo* (Kolonin, MG, and Finley, RL, Genetics, 1998 95:4266-4271). For example, transgenic 30 *Drosophila* lines expressing the desired aptamers may be generated by P element mediated transformation (discussed below). The phenotypes of the progeny expressing the aptamers can then be characterized.

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 BRCA2 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 the BRCA2 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 the BRCA2 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* BRCA2 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).

A preferred method of transposon mutagenesis in *Drosophila* employs the "local hopping" method described by Tower *et al.* (Genetics (1993) 133:347-359). Each new P insertion line can be tested molecularly for transposition of the P element into the gene of interest (*e.g.* BRCA2) by assays based on PCR. For each reaction, one PCR primer is used that is homologous to sequences contained within the P element and a second primer is homologous to the coding region or flanking regions of the gene of interest. Products of the PCR reactions are detected by agarose gel electrophoresis. The sizes of the resulting DNA fragments reveal the site of P element insertion relative to the gene of interest. Alternatively, Southern blotting and restriction mapping using DNA probes derived from genomic DNA or cDNAs of the gene of interest can be used to detect transposition events that rearrange the genomic DNA of the gene. P transposition events that map to the gene of interest can be assessed for phenotypic effects in heterozygous or homozygous mutant *Drosophila.*

In another embodiment, *Drosophila* lines carrying P insertions in the gene of interest, can be used to generate localized deletions using known methods (Kaiser, Bioassays (1990) 12(6):297-301; Harnessing the power of *Drosophila* genetics, In Drosophila melanogaster: Practical Uses in Cell and Molecular Biology, Goldstein and Fyrberg, Eds., Academic Press, Inc. San Diego, California). This is particularly useful if no P element transpositions are found that disrupt the gene of interest. Briefly, flies containing P elements inserted near the gene of interest are exposed to a further round of

transposase to induce excision of the element. Progeny in which the transposon has excised are typically identified by loss of the eye color marker associated with the transposable element. The resulting progeny will include flies with either precise or imprecise excision of the P element, where the imprecise excision events often result in deletion of genomic DNA neighboring the site of P insertion. Such progeny are screened by molecular techniques to identify deletion events that remove genomic sequence from the gene of interest, and assessed for phenotypic effects in heterozygous and homozygous mutant *Drosophila*.

Recently a transgenesis system has been described that may have universal applicability in all eye-bearing animals and which has been proven effective in delivering transgenes to diverse insect species (Berghammer *et al.*, Nature (1999) 402:370-371). This system includes: an artificial promoter active in eye tissue of all animal species, preferably containing three Pax6 binding sites positioned upstream of a TATA box (3xP3; Sheng *et al.*, Genes Devel. (1997) 11:1122-1131); a strong and visually detectable marker gene, such as GFP or other autofluorescent protein genes (Pasher *et al.*, Gene (1992) 111:229-233; U.S. Pat. No 5,491,084); and promiscuous vectors capable of delivering transgenes to a broad range of animal species. Examples of promiscuous vectors include transposon-based vectors derived from *Hermes*, *PiggyBac*, or *mariner*, and vectors based on pantropic VSV_G-pseudotyped retroviruses (Burns *et al.*, In Vitro Cell Dev Biol Anim (1996) 32:78-84; Jordan *et al.*, Insect Mol Biol (1998) 7: 215-222; U.S. Pat. No. 5,670,345). Thus, since the same transgenesis system can be used in a variety of phylogenetically diverse animals, comparative functional studies are greatly facilitated, which is especially helpful in evaluating new applications to pest management.

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 *hsp 16-2* and *hsp 16-41*. 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 *vestigal* 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 the BRCA2 pathway genes.

In *Drosophila*, binary control systems that employ exogenous DNA are useful when testing the mis-expression of genes in a wide variety of developmental stage-specific and tissue-specific patterns. Two examples of binary exogenous regulatory systems include the UAS/GAL4 system from yeast (Hay *et al.*, PNAS (1997) 94(10):5195-5200; Ellis *et al.*, Development (1993) 119(3):855-865), and the "Tet system" derived from *E. coli* (Bello et al., Development (1998) 125:2193-2202). The UAS/GAL4 system is a well-established and powerful method of mis-expression in *Drosophila* which employs the UAS_G upstream regulatory sequence for control of promoters by the yeast GAL4 transcriptional activator protein (Brand and Perrimon, Development (1993) 118(2):401-15). In this approach, transgenic *Drosophila*, termed "target" lines, are generated where the gene of interest to be mis-expressed is operably fused to an appropriate promoter controlled by UAS_G. Other transgenic *Drosophila* strains, termed "driver" lines, are generated where the GAL4 coding region is operably fused to promoters/enhancers that

direct the expression of the GAL4 activator protein in specific tissues, such as the eye, wing, nervous system, gut, or musculature. The gene of interest is not expressed in the target lines for lack of a transcriptional activator to drive transcription from the promoter joined to the gene of interest. However, when the UAS-target line is crossed with a GAL4 driver line, mis-expression of the gene of interest is induced in resulting progeny in a specific pattern that is characteristic for that GAL4 line. The technical simplicity of this approach makes it possible to sample the effects of directed mis-expression of the gene of interest in a wide variety of tissues by generating one transgenic target line with the gene of interest, and crossing that target line with a panel of pre-existing driver lines.

In the "Tet" binary control system, transgenic *Drosophila* driver lines are generated where the coding region for a tetracycline-controlled transcriptional activator (tTA) is operably fused to promoters/enhancers that direct the expression of tTA in a tissue-specific and/or developmental stage-specific manner. The driver lines are crossed with transgenic *Drosophila* target lines where the coding region for the gene of interest to be mis-expressed is operably fused to a promoter that possesses a tTA-responsive regulatory element. When the resulting progeny are supplied with food supplemented with a sufficient amount of tetracycline, expression of the gene of interest is blocked. Expression of the gene of interest can be induced at will simply by removal of tetracycline from the food. Also, the level of expression of the gene of interest can be adjusted by varying the level of tetracycline in the food. Thus, the use of the Tet system as a binary control mechanism for mis-expression has the advantage of providing a means to control the amplitude and timing of mis-expression of the gene of interest, in addition to spatial control. Consequently, if a gene of interest (*e.g.* a BRCA2 gene) has lethal or deleterious effects when mis-expressed at an early stage in development, such as the embryonic or larval stages, the function of the gene of interest in the adult can still be assessed by adding tetracycline to the food during early stages of development and removing tetracycline later so as to induce mis-expression only at the adult stage.

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-function phenotypes in the presence of a normal copy of the gene, can be made using known methods (Hershkowitz, Nature (1987) 329:219-222). In the case of active monomeric proteins, overexpression of an inactive form, achieved, for example, by linking the mutant gene to a highly active promoter, can cause competition for natural substrates or ligands sufficient to significantly reduce net activity of the normal

protein. Alternatively, changes to active site residues can be made to create a virtually irreversible association with a target.

Assays for Change in Gene Expression

5 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 BRCA2 gene compared to another cell line or animal expressing a mutant BRCA2 gene. Such expression profiling techniques include differential display, serial analysis of gene expression (SAGE), transcript profiling coupled to a gene database query, nucleic acid 10 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 BRCA2 gene. Gene expression profiling can also be used to identify other genes (or proteins) that may have a functional 15 relation to BRCA2 (*e.g.* may participate in a signaling pathway with the BRCA2 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 BRCA2 gene.

20 **Phenotypes Associated With BRCA2 Pathway Gene Mutations**

After isolation of model animals carrying mutated or mis-expressed BRCA2 pathway genes or inhibitory RNAs, animals are carefully examined for phenotypes of interest. For analysis of BRCA2 pathway genes that have been mutated (i.e. deletions, insertions, and/or point mutations) animal models that are both homozygous and 25 heterozygous for the altered BRCA2 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 pharmaceuticals. Some phenotypes more specific to flies include alterations in: adult behavior such as, flight ability, walking, grooming, phototaxis, mating or egg-30 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 discs; 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.

5 Genomic sequences containing a BRCA2 pathway gene can be used to confirm whether an existing mutant insect or worm line corresponds to a mutation in one or more BRCA2 pathway genes, by rescuing the mutant phenotype. Briefly, a genomic fragment containing the BRCA2 pathway gene of interest and potential flanking regulatory regions can be subcloned into any appropriate insect (such as *Drosophila*) transformation vector, 10 and injected into the animals. For *Drosophila*, an appropriate helper plasmid is used in the injections to supply transposase for transposon-based vectors. Resulting germline transformants are crossed for complementation testing to an existing or newly created panel of *Drosophila* lines whose mutations have been mapped to the vicinity of the gene of interest (Fly Pushing: The Theory and Practice of *Drosophila* Genetics, *supra*). If a 15 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 BRCA2 pathway gene. This prediction can be further confirmed by sequencing the BRCA2 pathway gene from the mutant line to identify the lesion in the BRCA2 pathway gene.

20

Identification of Genes That Modify BRCA2 Genes

The characterization of new phenotypes created by mutations or misexpression in BRCA2 genes enables one to test for genetic interactions between BRCA2 genes and other genes that may participate in the same, related, or interacting genetic or biochemical 25 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 BRCA2 genes with other well-characterized genes, particularly genes involved in cellular response to DNA damage.

30

Genetic Modifier Screens

A genetic modifier screen using invertebrate model organisms is a particularly preferred method for identifying genes that interact with BRCA2 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 BRCA2 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) 143:315-329). 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 BRCA2 genes using
5 methods described above.

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
10 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 BRCA2 modifier genes are identified, homologous genes in other species can be isolated using procedures based on cross-hybridization with modifier gene DNA
15 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 BRCA2 genes, human and rodent homologs of the modifier genes are of particular interest.

Although the above-described *Drosophila* genetic modifier screens are quite
20 powerful and sensitive, some genes that interact with BRCA2 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
25 *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
30 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 BRCA2 gene can be performed by crossing several thousand *Drosophila* EP lines individually into a genetic background containing a mutant or mis-expressed BRCA2 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 BRCA2 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 BRCA2 gene, have a physically tagged BRCA2 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, or 5, and how these sequences, and derivatives and fragments thereof, as well as other BRCA2 pathway nucleic acids and gene products can be used for genetic studies to elucidate mechanisms of the BRCA2 pathway as well as the discovery of potential pharmaceutical 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
5 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
10 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
15 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
20 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
25 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, dirhodamine or BigDye chemistries (Applied Biosystems, Inc., Foster City, CA).

Analysis of sequences were done as follows: the traces generated by the automated
30 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
5 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
10 sequences were archived in an Oracle-based relational database (FlyTag™, Exelixis Pharmaceuticals, Inc., South San Francisco, CA).

Example 2: Cloning of BRCA2 Nucleic Acid Sequences

Unless otherwise noted, the PCR conditions used for cloning the BRCA2 nucleic
15 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
20 analyzed and assembled into contigs using the Phred-Phrap programs.

For *Tribolium,* 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.
25 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. Skaletsky (1998) Primer3. Code available online at website www-
30 genome.wi.mit.edu/genome_software/other/primer3.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 a contiguous nucleotide sequence of 2.3 kilobases in length, encompassing an open reading frame (ORF) of 2271 nucleotides encoding a predicted protein of 756 amino acids. The ORF extends from base 1-2271 of SEQ ID NO:5.

For *Drosophila*, genomic sequencing of BAC clones led to a gene prediction for DBRCA2.2 (SEQ ID NO:3 for DNA and SEQ ID NO:4 for protein). DBRCA2.2 is an open reading frame of 2844 nucleotides, encoding a predicted protein of 947 amino acids. Sequences from the predicted gene DBRCA2.2 were used to probe and isolate the clone GM03728 from Berkeley *Drosophila* EST library (website at www.fruitfly.org). Sequencing of this clone led to a contiguous nucleotide sequence of 3.2 kilobases in length, encompassing an open reading frame (ORF) of 1920 nucleotides encoding a predicted protein of 640 amino acids. The ORF extends from base 1058-2977 of SEQ ID NO:1.

Example 3: Analysis of BRCA2 Nucleic Acid Sequences

Upon completion of cloning, the sequences were analyzed using the Pfam and Prosite programs. BRCA2 repeat domains were identified for each sequence. For DBRCA2.1 (SEQ ID NO:2), these repeats are at amino acids 239-273, 340-374, and 415-449. For DBRCA2.2 (SEQ ID NO: 4), these repeats are at amino acids 546-580, 647-681, and 722-756. For TBRCA2 (SEQ ID NO:6), the repeats are at amino acids 124-156, 158-192, and 198-232. Nucleotide and amino acid sequences for each of the BRCA2 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, where available, are listed for each BRCA2 gene.

TABLE 1

DNA BLAST, DBRCA2.1, SEQ ID NO:1

GI#	DESCRIPTION
7291775	AE003464 *Drosophila melanogaster* genomic scaffold 142000013386038 section 13 of 15, complete sequence
13162481	C007574 *Drosophila melanogaster*, chromosome 2R, region 60D-60D, BAC clone BACR11C07, complete sequence
DNA BLAST, DBRCA2.2, SEQ ID NO:3	
GI#	DESCRIPTION
7291775	AE003464 *Drosophila melanogaster* genomic scaffold 142000013386038 section 13 of 15, complete sequence
13162481	C007574 *Drosophila melanogaster*, chromosome 2R, region 60D-60D, BAC clone BACR11C07, complete sequence
DNA BLAST, TBRCA2, SEQ ID NO: 5	
GI#	DESCRIPTION
1513106	PCU65004 *Pneumocystis carinii* P-type proton motive membrane ATPase (PCA1) gene, complete cds
6434440	CEY39B6B *Caenorhabditis elegans* cosmid Y39B6B, complete sequence
15209321	CEY39B6A *Caenorhabditis elegans* cosmid Y39B6A, complete sequence
5824873	CEY79H2A *Caenorhabditis elegans* cosmid Y79H2A, complete sequence
12863250	AC087232 *Caenorhabditis elegans* cosmid Y92H12BR, complete sequence
PROTEIN BLAST, DBRCA2.1, SEQ ID NO:2	
GI#	DESCRIPTION
7291801	(AE003464) CG13584 gene product [*Drosophila melanogaster*]
1523876	(Z75666) BRCA2 [*Cercopithecus aethiops*]
13540359	(AF348515) mutant early onset breast cancer susceptibility protein 2 [*Homo sapiens*]
11277004	breast cancer tumor suppressor BRCA2 - dog (fragment)
1929048	(Z74739) 214K23.1 [*Homo sapiens*]
PROTEIN BLAST, DBRCA2.2, SEQ ID NO:4	
GI#	DESCRIPTION
7291800	(AE003464) CG13583 gene product [*Drosophila melanogaster*]
7291801	(AE003464) CG13584 gene product [*Drosophila melanogaster*]
1523876	(Z75666) BRCA2 [*Cercopithecus aethiops*]
13540359	(AF348515) mutant early onset breast cancer susceptibility protein 2 [*Homo sapiens*]
11277004	breast cancer tumor suppressor BRCA2 - dog (fragment)
PROTEIN BLAST, TBRCA2, SEQ ID NO:6	
GI#	DESCRIPTION
6857765	breast cancer 2 [*Mus musculus*]
1177438	(X95152) brca2 [*Homo sapiens*]
13928732	breast cancer 2 [*Rattus norvegicus*]
14757036	breast cancer 2, early onset [*Homo sapiens*]
4502451	breast cancer 2, early onset [*Homo sapiens*]

The closest homolog predicted by BLAST analysis for each sequence is a mammalian BRCA2.

Example 4: Cell survival and cell cycle checkpoint analysis

Cell clones are generated that express a GFP-BRC4 fusion protein, employeing the tetracycline-inducible expression system controlled by a tet-responsive promoter (Gossen M and Bujard H (1992) PNAS 89:5547-5551).

5 For cell survival assays, cells are seeded in identical plates at 5000 cells/plate in medium with tetracycline (1 μg/ml). Expression of the wild-type or mutant GFP-BRC4 repeat is induced by removing tetracycline 24 h after seeding. Twenty four hours after induction of GFP-BRC4 expression, cells are gamma-irradiated with 3 Gy. After incubation for 14 days, cells are fixed and stained with 2% methylene blue in 50% of
10 ethanol for colony counting. Averages and standard deviations are determined from eight plates.

For cell cycle checkpoint analysis, cells in logarithmic growth are mock-exposed or -irradiated (12 Gy). After 24 h, cells are labeled with 10 μM BrdUrd for 4 h and fixed for BrdUrd staining using a Cell Proliferation Kit (Amersham Pharmacia Biotech).
15 BrdUrd-positive cells are quantified, and expressed as a fraction of the total cells. For the G2/M checkpoint, cells are irradiated to 3 Gy, fixed with 4% paraformaldehyde at indicated time and stained with DAPI for counting mitotic cells in prophase, metaphase, anaphase, and telophase. Alternately, cells are irradiated with 4-16 Gy and processed for analysis of mitotic cells after 1 hr.

20

Example 5: BRCA2 expression patterns in Drosophila

BRCA2 expression levels in cDNA libraries and cell lines derived from *Drosophila* were assessed by TaqMan (Perkin Elmer Corp./Applied Biosystems, Foster City, Calif) Quantitative Reverse-transcriptase PCR (RT-PCR) (Livak KJ, et al., PCR
25 Methods Appl. 1995 Jun;4(6):357-62). Briefly, reverse transcribed cDNA from total RNA was produced from each S2 and Dmel2 *Drosophila* cell lines (S2 is available from American Type Culture Collection (ATCC), Cat#CRL-1963; Dmel 2 is available from Life Technologies, Cat#10831). Embryonic, adult, and larval cDNA *Drosophila* libraries were obtained from Berkeley *Drosophila* Genome Project (BDGP, website at
30 www.fruitfly.org). Primers were made using SEQ ID NO:1. Procedures for primer design and PCR conditions were according to manufacturer's protocols.

The results of these experiments indicate that BRCA2 has high expression levels in embryonic and adult cDNA libraries, medium expression levels in S2 and Dmel2 cells, and low expression levels in larval cDNA libraries.

Example 6: BRCA2 RNAi in Dmel2 cells

RNAi was used to knock down levels of BRCA2 in Dmel2 *Drosophila* cells. Briefly, PCR primers based on BRCA2 sequences of SEQ ID NO:1 were employed to
5 produce PCR products of about 1 kb, using an embryonic *Drosophila* cDNA library (Example 5) as template. In vitro transcription (Ambion, TX) was then used to produce single stranded (ss) RNA, which was then annealed to produce double stranded (ds) RNA. Cultured *Drosophila* embryonic serum-free medium adapted Dmel2 cell line was treated with BRCA2 dsRNA and allowed to incubate for 2 days. RNA was then extracted from
10 the untreated (control) and treated cells (Qiagen, Valencia, CA), and reverse transcribed to produce cDNA (ABI kit, Applied Biosystems, Foster City, CA). The reverse transcribed cDNA was used as a template for RT-PCR using Taqman experiments. Taqman primers and procedures were produced and performed according to manufacturer's suggestions. The results show that BRCA2 RNAi in Dmel2 cells reduces BRCA2 mRNA levels by
15 80%.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule of less than about 15 kb in size comprising a nucleic acid sequence selected from the group consisting of:

5 (a) a nucleic acid sequence that encodes a polypeptide comprising amino acids of SEQ ID NOs:2, 4, or 6; and

(b) the complement of the nucleic acid sequence of (a).

2. The isolated nucleic acid molecule of Claim 1 which is capable of hybridizing to a 10 nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, or 5 using highly stringent hybridization conditions.

3. The isolated nucleic acid molecule of Claim 1 wherein said nucleic acid sequence encodes the entire sequence of SEQ ID NOs:2, 4, or 6.

15

4. The isolated nucleic acid molecule of Claim 1 wherein said nucleic acid sequence encodes at least one BRC repeat of the BRCA2.

5. A vector comprising the nucleic acid molecule of Claim 1.

20

6. An isolated host cell comprising the vector of Claim 7.

7. A process for producing a BRCA2 protein comprising culturing the host cell of Claim 8 under conditions suitable for expression of said BRCA2 protein and recovering 25 said protein.

8. A method for detecting a candidate compound that interacts with a BRCA2 protein or fragment thereof, said method comprising contacting said BRCA2 protein or fragment with one or more candidate molecules, and detecting any interaction between said 30 candidate compound and said BRCA2 protein or fragment; wherein the amino acid sequence of said BRCA2 protein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:2, 4, or 6.

9. The method of Claim 8 wherein said candidate compound is a putative pharmaceutical agent.

10. The method of Claim 8 wherein said contacting comprises administering said candidate compound to cultured host cells that have been genetically engineered to express said BRCA2 protein.

11. The method of Claim 8 wherein said contacting comprises administering said candidate compound to a metazoan invertebrate organism that has been genetically engineered to express a BRCA2 protein.

12. A first animal that is an insect that has been genetically modified to express or mis-express a BRCA2 protein, or the progeny of said animal that has inherited said BRCA2 protein expression or mis-expression, wherein said BRCA2 protein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:2, 4, or 6.

13. A method for studying BRCA2 activity comprising detecting the phenotype caused by the expression or mis-expression of said BRCA2 protein in the first animal of Claim 12.

14. The method of Claim 13 additionally comprising observing a second animal having the same genetic modification as said first animal which causes said expression or mis-expression of said BRCA2 protein, and wherein said second animal additionally comprises a mutation in a gene of interest, wherein differences, if any, between the phenotype of said first animal and the phenotype of said second animal identifies the gene of interest as capable of modifying the function of the gene encoding said BRCA2 protein.

15. The method of Claim 13 additionally comprising administering one or more candidate compounds to said animal or its progeny and observing any changes in BRCA2 activity of said animal or its progeny.

SEQUENCE LISTING

<110> Exelixis, Inc.

<120> NUCLEIC ACIDS AND POLYPEPTIDES OF INVERTEBRATE BRCA2 AND METHODS OF USE

<130> EX02-123C-PC

<150> 60/332,947
<151> 2001-11-16

<160> 6

<170> PatentIn version 3.1

<210> 1
<211> 3181
<212> DNA
<213> Drosophila melanogaster

<400> 1
aaccacaaaa gcattggaca cttgcttact tcaaataaaa cacatttaaa caattagatg 60

acgtgatgga ccagaatggc gcatcgggaa gtcatccgaa caggctatcc caaggaagag 120

gagcccatgc gcgcgaacgt ggcgccacag tttccgcggc ggcaaatcgg agtaacatta 180

tcgacgaaat ggccaaaata tgcgaagccg atcgccagac tttcgccatc gctcgacgga 240

ctcggggtca cgagcggctt gcggtggaca acagcgactt cgtcgccgtg gaggatctta 300

ttttgtccta cgcagagccc acgcccgagg accaggtcga gatgatcatg agcgactttt 360

gctcgtctcc aacatacgca gaggatgagg atgagcccag ccatgagtcg gagccgtggt 420

ttcgatttcg caacaaaagg atcagaacct acagccggaa gagggatccc aaaagccaca 480

aggccgttca aaacgagaag cgtagaggtt cctcaggcct ctccgtgcag agggatctca 540

atacttcgtt cacatctatg gcttgtgatt tcgatgcttc atcacagaag atacacgagg 600

tccttttgaa cctcagtcaa tacttttccg cgaccgcgac agcttccggt ccgactcctg 660

tcccatcgca aatagatctg ccaaccgaag caaggcagga ttctggtgac gagtgtttca 720

acgctgaatt ggagaacttg agggatgacc tgttgcagag cggatttaca tttgaggcct 780

ctttctatga aaatgagcat gagcaggaag gtccaggaac tatggcagac actatttatg 840

gcgatagcat gcaatcggtt cccactacgc gcctccgatt ggaaagtgga aaacgtaatg 900

tttggtctga tgcagactct acattgaaaa aagctgatgt cgaaattgct gaaaataaga 960

aacttttgat gggccaaacg caagcaaaaa tgttgacata gaggagaaca caaacttcat 1020

attagaaggc ataccactaa gtgaatggct aactccgatg gaactgcccg agatctccaa 1080

agatgtaatc aaacatatac ccgataaaaa gttaaagttg gagccaagtt cccagaaaga 1140

```
acaaaaaagt agcaaagatt caaatgcatc aaaggttaga ggagcctcaa aacaatcttg    1200
tgatattaac acgaaaaacg aaggaactac gattctggat cagcctaatg ctgcagagca    1260
agaaaaccta ttgaacgacg gcgagcttct tgaggagttt ctttttaatg attggcaacc    1320
aatgcagtgc tctaatggcc catcaacctc caaaaacgca attcaagggc ctaaagaaaa    1380
catcaattca ataaatctgg acgacaagga gcaaccagaa aagcagactc ctaataaaag    1440
tcagaccata agcagtcatc aacttaatgg cattcgaaag aagtccttca aatttatcga    1500
ggtttcggaa gaaatgaaaa ttaaaggcga agagtttgta gacaaagttg tatccggcct    1560
atatcactct agtcacaagt gcaacttacg tactgaggaa tactccgata atcatagtca    1620
agtaatggaa tcaacacagt gtgcggaatt caaatcagcc ccaagcaagc ccatcgaaat    1680
attggatggt aaggaaacct acaatgcaat agctaaagtt gacgttggag aaataaatgg    1740
aaaattcagt cctctcaata acgagactat agcagagccc gagttttgtg gctttcgcac    1800
agcctcaaat aaggcaattc cgatttctga gaagatgaaa atcaaaactg cagagtttat    1860
ggctgagttt caatccaaag aaacaattca acaaaatgat tatttagtcc atcagccaaa    1920
tgacaaaccg acgtctgttg gacttgatac agccttaaaa agatcaatag agagttcaga    1980
ggaaatgcga tcgaaagctt ccaagttggt ggacgttgat actacaatgc gggaaccgca    2040
tcaacccaca cttgacccag tttgtcgtga cttaaatgaa tcacagtttt tcggctttag    2100
aacagcgtca aacaaagcca ttgagattac cgaggcgatg gagaagagag gagccatgtt    2160
tctggcccag tctaaagcta cagaccaatt aaacgggtgg cagccaagcg attttcctga    2220
tgttctaccc acctctccaa acaatgagat tcattcaata aatgttgaaa ataataaggc    2280
tgtgaatact aaaaccgtat ctgagaccga attcttcggc tttcgtaccg cttcaaataa    2340
aggaattgtg atttcggaaa atacgaaaat caaagttgca cagttcatga gtgaatttca    2400
agccgccgat gcaagcactg actcgaacca accaattgtg atctctgagg aatcgagaaa    2460
tattgatgcc aagttcgtag atgaagcggc ggcagaagac agtgccaaca agccaacttt    2520
ttgcaatgta caaagtctag aaaatacctc agatattgag catttcaaac acgacttctt    2580
cgtggagcat tctgcaaaag aagagcatcc gctgtgtagc cagccattgg tgaggacacc    2640
cagaaggtcg caggagatcc actcctcact atcgcaactt gccggcaaat caccactgga    2700
ccaggctacc aagaagtccg taatagctcg ccggaacctc ctgtcactaa aaagaaagcg    2760
aaagatcatt tcgtcaactg aaacttccac ttcctgcgcc ttacccacta tggagcgatt    2820
cgctcccaag ccttcctcga caagtactcc gctggccgac agggacttaa atcgctccaa    2880
ggactgtaca aagaatagac aggatgcgga ggacatgtcc cctatttgca tgcaacccaa    2940
```

aaagagtcga agacttggtc tgagtcgaag tcgctattag tacacattaa gtaagcgttg 3000

ctttataccc aaaatattta aaccgctggc cgatttgaaa tgggtagaga tttagttgaa 3060

ttaattaagg tatgccatta gtggtctcat catgacagat tgaataattg tataagttat 3120

cataataagc gttttttgtt aagtgcacct tataaggaaa taaacgatcc gataagggcc 3180

a 3181

<210> 2
<211> 640
<212> PRT
<213> Drosophila melanogaster

<400> 2

Met Glu Leu Pro Glu Ile Ser Lys Asp Val Ile Lys His Ile Pro Asp
1 5 10 15

Lys Lys Leu Lys Leu Glu Pro Ser Ser Gln Lys Glu Gln Lys Ser Ser
20 25 30

Lys Asp Ser Asn Ala Ser Lys Val Arg Gly Ala Ser Lys Gln Ser Cys
35 40 45

Asp Ile Asn Thr Lys Asn Glu Gly Thr Thr Ile Leu Asp Gln Pro Asn
50 55 60

Ala Ala Glu Gln Glu Asn Leu Leu Asn Asp Gly Glu Leu Leu Glu Glu
65 70 75 80

Phe Leu Phe Asn Asp Trp Gln Pro Met Gln Cys Ser Asn Gly Pro Ser
85 90 95

Thr Ser Lys Asn Ala Ile Gln Gly Pro Lys Glu Asn Ile Asn Ser Ile
100 105 110

Asn Leu Asp Asp Lys Glu Gln Pro Glu Lys Gln Thr Pro Asn Lys Ser
115 120 125

Gln Thr Ile Ser Ser His Gln Leu Asn Gly Ile Arg Lys Lys Ser Phe
130 135 140

Lys Phe Ile Glu Val Ser Glu Glu Met Lys Ile Lys Gly Glu Glu Phe
145 150 155 160

Val Asp Lys Val Val Ser Gly Leu Tyr His Ser Ser His Lys Cys Asn
165 170 175

```
Leu Arg Thr Glu Glu Tyr Ser Asp Asn His Ser Gln Val Met Glu Ser
            180                 185                 190

Thr Gln Cys Ala Glu Phe Lys Ser Ala Pro Ser Lys Pro Ile Glu Ile
        195                 200                 205

Leu Asp Gly Lys Glu Thr Tyr Asn Ala Ile Ala Lys Val Asp Val Gly
    210                 215                 220

Glu Ile Asn Gly Lys Phe Ser Pro Leu Asn Asn Glu Thr Ile Ala Glu
225                 230                 235                 240

Pro Glu Phe Cys Gly Phe Arg Thr Ala Ser Asn Lys Ala Ile Pro Ile
                245                 250                 255

Ser Glu Lys Met Lys Ile Lys Thr Ala Glu Phe Met Ala Glu Phe Gln
            260                 265                 270

Ser Lys Glu Thr Ile Gln Gln Asn Asp Tyr Leu Val His Gln Pro Asn
        275                 280                 285

Asp Lys Pro Thr Ser Val Gly Leu Asp Thr Ala Leu Lys Arg Ser Ile
    290                 295                 300

Glu Ser Ser Glu Glu Met Arg Ser Lys Ala Ser Lys Leu Val Asp Val
305                 310                 315                 320

Asp Thr Thr Met Arg Glu Pro His Gln Pro Thr Leu Asp Pro Val Cys
                325                 330                 335

Arg Asp Leu Asn Glu Ser Gln Phe Phe Gly Phe Arg Thr Ala Ser Asn
            340                 345                 350

Lys Ala Ile Glu Ile Thr Glu Ala Met Glu Lys Arg Gly Ala Met Phe
        355                 360                 365

Leu Ala Gln Ser Lys Ala Thr Asp Gln Leu Asn Gly Trp Gln Pro Ser
    370                 375                 380

Asp Phe Pro Asp Val Leu Pro Thr Ser Pro Asn Asn Glu Ile His Ser
385                 390                 395                 400

Ile Asn Val Glu Asn Asn Lys Ala Val Asn Thr Lys Thr Val Ser Glu
                405                 410                 415
```

```
Thr Glu Phe Phe Gly Phe Arg Thr Ala Ser Asn Lys Gly Ile Val Ile
            420                 425                 430

Ser Glu Asn Thr Lys Ile Lys Val Ala Gln Phe Met Ser Glu Phe Gln
        435                 440                 445

Ala Ala Asp Ala Ser Thr Asp Ser Asn Gln Pro Ile Val Ile Ser Glu
    450                 455                 460

Glu Ser Arg Asn Ile Asp Ala Lys Phe Val Asp Glu Ala Ala Ala Glu
465                 470                 475                 480

Asp Ser Ala Asn Lys Pro Thr Phe Cys Asn Val Gln Ser Leu Glu Asn
                485                 490                 495

Thr Ser Asp Ile Glu His Phe Lys His Asp Phe Phe Val Glu His Ser
            500                 505                 510

Ala Lys Glu Glu His Pro Leu Cys Ser Gln Pro Leu Val Arg Thr Pro
        515                 520                 525

Arg Arg Ser Gln Glu Ile His Ser Ser Leu Ser Gln Leu Ala Gly Lys
    530                 535                 540

Ser Pro Leu Asp Gln Ala Thr Lys Lys Ser Val Ile Ala Arg Arg Asn
545                 550                 555                 560

Leu Leu Ser Leu Lys Arg Lys Arg Lys Ile Ile Ser Ser Thr Glu Thr
                565                 570                 575

Ser Thr Ser Cys Ala Leu Pro Thr Met Glu Arg Phe Ala Pro Lys Pro
            580                 585                 590

Ser Ser Thr Ser Thr Pro Leu Ala Asp Arg Asp Leu Asn Arg Ser Lys
        595                 600                 605

Asp Cys Thr Lys Asn Arg Gln Asp Ala Glu Asp Met Ser Pro Ile Cys
    610                 615                 620

Met Gln Pro Lys Lys Ser Arg Arg Leu Gly Leu Ser Arg Ser Arg Tyr
625                 630                 635                 640

<210>  3
<211>  2844
```

<212> DNA
<213> Drosophila melanogaster

<400> 3
atggaccaga atggcgcatc gggaagtcat ccgaacaggc tatcccaagg aagaggagcc 60

catgcgcgcg aacgtggcgc cacagtttcc gcggcggcaa atcggagtaa cattatcgac 120

gaaatggcca aaatatgcga agccgatcgc cagactttcg ccatcgctcg acggactcgg 180

ggtcacgagc ggcttgcggt ggacaacagc gacttcgtcg ccgtggagga tcttattttg 240

tcctacgcag agcccacgcc cgaggaccag gtcgagatga tcatgagcga cttttgctcg 300

tctccaacat acgcagagga tgaggatgag cccagccatg agtcggagcc gtggtttcga 360

tttcgcaaca aaaggatcag aacctacagc cggaagaggg atcccaaaag ccacaaggcc 420

gttcaaaacg agaagcgtag aggttcctca ggcctctccg tgcagaggga tctcaatact 480

tcgttcacat ctatggcttg tgatttcgat gcttcatcac agaagataca cgaggtcctt 540

ttgaacctca gtcaatactt ttccgcgacc gcgacagctt ccggtccgac tcctgtccca 600

tcgcaaatag atctgccaac cgaagcaagg caggattctg gtgacgagtg tttcaacgct 660

gaattggaga acttgaggga tgacctgttg cagagcggat ttacatttga ggcctctttc 720

tatgaaaatg agcatgagca ggaagaaagt ggaaaacgta atgtttggtc tgatgcagac 780

tctacattga aaaaagctga tgtcgaaatt gctgaaaata agaaactttt gatgggccaa 840

acgcaagcaa aaaatgttga catagaggag aacacaaact tcatattaga aggcatacca 900

ctaagtgaat ggctaactcc gatggaactg cccgagatct ccaaagatgt aatcaaacat 960

atacccgata aaaagttaaa gttggagcca agttcccaga aagaacaaaa aagtagcaaa 1020

gattcaaatg catcaaaggt tagaggagcc tcaaaacaat cttgtgatat taacacgaaa 1080

aacgaaggaa ctactattct ggatcagcct aatgctgcag agcaagaaaa cctattgaac 1140

gacggcgagc ttcttgagga gtttcttttt aatgattggc aaccaatgca gtgctctaat 1200

ggcccatcaa cctccaaaaa cgcaattcaa gggcctaaag aaaacatcaa ttcaataaat 1260

ctggacgaca aggagcaacc agaaaagcag actcctaata aaagtcagac cataagcagt 1320

catcaactta atggcattcg aaagaagtcc ttcaaattta tcgaggtttc ggaagaaatg 1380

aaaattaaag gcgaagagtt tgtagacaaa gttgtatccg gcctatatca ctttagtcac 1440

aagtgcaact tacgtactga ggaatactcc gataatcata gtcaagtaat ggaatcaaca 1500

cagtgtgcgg aattcaaatc agccccaagc aagcccatcg aaatattgga tggtaaggaa 1560

acctacaatg caatagctaa agttgacgtt ggagaaataa atggaaaatt cagtcctctc 1620

aataacgata ctatagcaga gcccgagttt tgtggctttc gcacagcctc aaataaggca 1680

```
attccgattt ctgagaagat gaaaatcaaa actgcagagt ttatggctga gtttcaatcc   1740
aaagaaacaa ttcaacaaaa tgattattta gtccatcagc caaatgacaa accgacgtct   1800
gttggacttg atacagcctt aaaaagatca atagagagtt cagaggaaat gcgatcgaaa   1860
gcttccaagt tggtggtcgt tgatactaca atgcgggaac cgcatcaacc cacacttgac   1920
ccagtttgcc gtgacttaaa tgaatcacag tttttcggct ttagaacagc gtcaaacaaa   1980
gccattgaga ttaccgaggc gatggagaag agaggagcca tgtttctggc ccagtctaaa   2040
gctacagacc aattaaacgg gtggcagcca agcgattttc ctgatgttct acccacctct   2100
ccaaacaatg agattcattc aataaatgtt gaaaataata aggctgtgaa tactaaaacc   2160
gtatctgaga ccgaattctt cggctttcgt accgcttcaa ataaaggaat tgtgatttcg   2220
gaaaatacga aaatcaaagt tgcacagttc atgagtgaat ttcaagccgc cgatgcaagc   2280
actgactcga accaaccaac tgtgatctct gaggaatcga gaaatattga tgccaagttc   2340
gtagatgaag cggcggcaga agacagtgcc aacaagccaa ctttttgcaa tgtacaaagt   2400
ctaaaaaata cctcagatat tgagcatttc aaacacgact tcttcgtgga gcattctgca   2460
aaggaagagc atccgctgtg tagccagcca ttggtgagga cacccagaag gtcgcaggag   2520
atccactcct cactatcgca acttgccggc aaatcaccac tggaccaggc caccaagaag   2580
tccgtaatag ctcgccggaa cctcctgtca ctaaaaagaa agcgaaagat cattttgtca   2640
actgaaactt ccacttcctg cgccttaccc actatggagc gattcgctcc caagccttcc   2700
tcgacaagta ctccgctggc cgacagggac ttaaatcgct ccaaggactg tacaaagaat   2760
agacaggatg cggaggacat gtcccctatt tgcatgcaac ccaaaaagag tcgaagactt   2820
ggtctgagtc gaagtcgcta ttag                                          2844
```

<210> 4
<211> 947
<212> PRT
<213> Drosophila melanogaster

<400> 4

```
Met Asp Gln Asn Gly Ala Ser Gly Ser His Pro Asn Arg Leu Ser Gln
1               5                   10                  15

Gly Arg Gly Ala His Ala Arg Glu Arg Gly Ala Thr Val Ser Ala Ala
            20                  25                  30

Ala Asn Arg Ser Asn Ile Ile Asp Glu Met Ala Lys Ile Cys Glu Ala
        35                  40                  45
```

```
Asp Arg Gln Thr Phe Ala Ile Ala Arg Arg Thr Arg Gly His Glu Arg
    50                  55                  60

Leu Ala Val Asp Asn Ser Asp Phe Val Ala Val Glu Asp Leu Ile Leu
65                  70                  75                  80

Ser Tyr Ala Glu Pro Thr Pro Glu Asp Gln Val Glu Met Ile Met Ser
                85                  90                  95

Asp Phe Cys Ser Ser Pro Thr Tyr Ala Glu Asp Glu Asp Glu Pro Ser
            100                 105                 110

His Glu Ser Glu Pro Trp Phe Arg Phe Arg Asn Lys Arg Ile Arg Thr
        115                 120                 125

Tyr Ser Arg Lys Arg Asp Pro Lys Ser His Lys Ala Val Gln Asn Glu
    130                 135                 140

Lys Arg Arg Gly Ser Ser Gly Leu Ser Val Gln Arg Asp Leu Asn Thr
145                 150                 155                 160

Ser Phe Thr Ser Met Ala Cys Asp Phe Asp Ala Ser Ser Gln Lys Ile
                165                 170                 175

His Glu Val Leu Leu Asn Leu Ser Gln Tyr Phe Ser Ala Thr Ala Thr
            180                 185                 190

Ala Ser Gly Pro Thr Pro Val Pro Ser Gln Ile Asp Leu Pro Thr Glu
        195                 200                 205

Ala Arg Gln Asp Ser Gly Asp Glu Cys Phe Asn Ala Glu Leu Glu Asn
    210                 215                 220

Leu Arg Asp Asp Leu Leu Gln Ser Gly Phe Thr Phe Glu Ala Ser Phe
225                 230                 235                 240

Tyr Glu Asn Glu His Glu Gln Glu Glu Ser Gly Lys Arg Asn Val Trp
                245                 250                 255

Ser Asp Ala Asp Ser Thr Leu Lys Lys Ala Asp Val Glu Ile Ala Glu
            260                 265                 270

Asn Lys Lys Leu Leu Met Gly Gln Thr Gln Ala Lys Asn Val Asp Ile
        275                 280                 285
```

```
Glu Glu Asn Thr Asn Phe Ile Leu Glu Gly Ile Pro Leu Ser Glu Trp
    290                 295                 300

Leu Thr Pro Met Glu Leu Pro Glu Ile Ser Lys Asp Val Ile Lys His
305                 310                 315                 320

Ile Pro Asp Lys Lys Leu Lys Leu Glu Pro Ser Ser Gln Lys Glu Gln
                325                 330                 335

Lys Ser Ser Lys Asp Ser Asn Ala Ser Lys Val Arg Gly Ala Ser Lys
            340                 345                 350

Gln Ser Cys Asp Ile Asn Thr Lys Asn Glu Gly Thr Thr Ile Leu Asp
        355                 360                 365

Gln Pro Asn Ala Ala Glu Gln Glu Asn Leu Leu Asn Asp Gly Glu Leu
    370                 375                 380

Leu Glu Glu Phe Leu Phe Asn Asp Trp Gln Pro Met Gln Cys Ser Asn
385                 390                 395                 400

Gly Pro Ser Thr Ser Lys Asn Ala Ile Gln Gly Pro Lys Glu Asn Ile
                405                 410                 415

Asn Ser Ile Asn Leu Asp Asp Lys Glu Gln Pro Glu Lys Gln Thr Pro
            420                 425                 430

Asn Lys Ser Gln Thr Ile Ser Ser His Gln Leu Asn Gly Ile Arg Lys
        435                 440                 445

Lys Ser Phe Lys Phe Ile Glu Val Ser Glu Glu Met Lys Ile Lys Gly
    450                 455                 460

Glu Glu Phe Val Asp Lys Val Val Ser Gly Leu Tyr His Phe Ser His
465                 470                 475                 480

Lys Cys Asn Leu Arg Thr Glu Glu Tyr Ser Asp Asn His Ser Gln Val
                485                 490                 495

Met Glu Ser Thr Gln Cys Ala Glu Phe Lys Ser Ala Pro Ser Lys Pro
            500                 505                 510

Ile Glu Ile Leu Asp Gly Lys Glu Thr Tyr Asn Ala Ile Ala Lys Val
        515                 520                 525
```

```
Asp Val Gly Glu Ile Asn Gly Lys Phe Ser Pro Leu Asn Asn Asp Thr
    530                 535                 540

Ile Ala Glu Pro Glu Phe Cys Gly Phe Arg Thr Ala Ser Asn Lys Ala
545                 550                 555                 560

Ile Pro Ile Ser Glu Lys Met Lys Ile Lys Thr Ala Glu Phe Met Ala
                565                 570                 575

Glu Phe Gln Ser Lys Glu Thr Ile Gln Gln Asn Asp Tyr Leu Val His
            580                 585                 590

Gln Pro Asn Asp Lys Pro Thr Ser Val Gly Leu Asp Thr Ala Leu Lys
        595                 600                 605

Arg Ser Ile Glu Ser Ser Glu Glu Met Arg Ser Lys Ala Ser Lys Leu
    610                 615                 620

Val Val Val Asp Thr Thr Met Arg Glu Pro His Gln Pro Thr Leu Asp
625                 630                 635                 640

Pro Val Cys Arg Asp Leu Asn Glu Ser Gln Phe Phe Gly Phe Arg Thr
                645                 650                 655

Ala Ser Asn Lys Ala Ile Glu Ile Thr Glu Ala Met Glu Lys Arg Gly
            660                 665                 670

Ala Met Phe Leu Ala Gln Ser Lys Ala Thr Asp Gln Leu Asn Gly Trp
        675                 680                 685

Gln Pro Ser Asp Phe Pro Asp Val Leu Pro Thr Ser Pro Asn Asn Glu
    690                 695                 700

Ile His Ser Ile Asn Val Glu Asn Asn Lys Ala Val Asn Thr Lys Thr
705                 710                 715                 720

Val Ser Glu Thr Glu Phe Phe Gly Phe Arg Thr Ala Ser Asn Lys Gly
                725                 730                 735

Ile Val Ile Ser Glu Asn Thr Lys Ile Lys Val Ala Gln Phe Met Ser
            740                 745                 750

Glu Phe Gln Ala Ala Asp Ala Ser Thr Asp Ser Asn Gln Pro Thr Val
        755                 760                 765
```

```
Ile Ser Glu Glu Ser Arg Asn Ile Asp Ala Lys Phe Val Asp Glu Ala
    770                 775                 780

Ala Ala Glu Asp Ser Ala Asn Lys Pro Thr Phe Cys Asn Val Gln Ser
785                 790                 795                 800

Leu Lys Asn Thr Ser Asp Ile Glu His Phe Lys His Asp Phe Phe Val
                805                 810                 815

Glu His Ser Ala Lys Glu Glu His Pro Leu Cys Ser Gln Pro Leu Val
            820                 825                 830

Arg Thr Pro Arg Arg Ser Gln Glu Ile His Ser Ser Leu Ser Gln Leu
        835                 840                 845

Ala Gly Lys Ser Pro Leu Asp Gln Ala Thr Lys Lys Ser Val Ile Ala
    850                 855                 860

Arg Arg Asn Leu Leu Ser Leu Lys Arg Lys Arg Lys Ile Ile Leu Ser
865                 870                 875                 880

Thr Glu Thr Ser Thr Ser Cys Ala Leu Pro Thr Met Glu Arg Phe Ala
                885                 890                 895

Pro Lys Pro Ser Ser Thr Ser Thr Pro Leu Ala Asp Arg Asp Leu Asn
            900                 905                 910

Arg Ser Lys Asp Cys Thr Lys Asn Arg Gln Asp Ala Glu Asp Met Ser
        915                 920                 925

Pro Ile Cys Met Gln Pro Lys Lys Ser Arg Arg Leu Gly Leu Ser Arg
    930                 935                 940

Ser Arg Tyr
945


<210>  5
<211>  2271
<212>  DNA
<213>  Tribolium castaneum

<400>  5
atggaagacc ggcctttatc acctattcta agcactcaaa tatcacgcaa atggattaga      60

agcaagagat cacagaaaaa aagtcgggtt aatttgacga accggtttct ggacgcgtct     120

tccagcaatg ttttaacccc aagccaaacc gaaaactttt tggaaagtat agaaaaagac     180
```

```
tggagtttgg ataaaaattc tgatgatgtg ttgaaaagag ccgaaatttt gactcctccc    240
caaggtttta attgtaaaaa tgaacccaac aaagttgata atcaaggttt caaagcgtct    300
gatgaagcct tgaaaaaggc cagaacgttg tttgacgaca tgctaaatcc tgaaggttgt    360
aacaaggaat taagtactgg ttttgcaggt tttaaaacag cttcaggaaa aaattcctta    420
agatccgaaa atgcagtcaa agaagctagg aataggaaca gtaatgaggg aactaccaca    480
agttttgtcg gttttaaaac agcttcaggg aaaagttccc taatgtctaa gagtgcaatc    540
aaacaagcta ggaataaatt aagcgaagat tttaatacga aaatttcgga atataacgca    600
agtttttcgg gttttaagac agcttcaggg aaatgtacag taatgtccga aagagcaatc    660
aaccaggcca ggaataaatt aaacgagaat tttaattttg aagacgttaa aaatgcgaaa    720
aaatcagatt tgagtgacga gttcagaggt tttagtcaag atatgattga gaaaaatcgg    780
aaaaaacttg acgaatttcg ctcgattgta acacaattta acaaaggtgt aaaaaaagaa    840
aaaaacttag aaaccgttcg ggaatttcct aaaaataggg gcttccaatc agtgaattta    900
tcaaaagaca ggctggaaaa agcgcgtcaa ttgttccaag gtctggacga ttcgtttggt    960
cgcaaaaggc gtttttttga gcaagaaaaa tttggtttct ctcctgttag aaaaaatccg   1020
agtgtttaca actcaacgcc tttgaaaaac acctctttta cgtgtctaga atcggacgta   1080
acaccgatca aaaaaatcaa aaacgagaca attattacac aaaatgatag atttgtaatt   1140
cagagctgca ctaaaggaaa tctcgatacg tggcttcaaa atttgcaaaa cgaacgaaaa   1200
attcttgaag taaaacttaa agttatcatc gagaaggaga aagttttaaa tttgcaaaaa   1260
gaaattctgg aaaatagggt tcagagacag tctggaattt tatttactag aaaagaaaat   1320
agcgaacgtg ttttgttaaa acatggcata aacatcgatt cgtgtcgtgt ttcaagtgaa   1380
attaattcgg aaaattcgcc caatgttcat ttgaagtacg tatgggtgtg ggaaaaactg   1440
ccaatgatta aaacgttaga tggtgcttgt gttgttccaa atgtcgataa tttcgttggt   1500
ttaagtgaga ttgaagtagc gtttaataca ataccaggag tcgaaaccaa attaatacca   1560
aatggttggg tgcaaaatca tttcagatgg attgtttgga agttggcgtg ttacgaaaat   1620
tttaatttcg attgtttaac agtcgaaaat attatgcaac aattgaaata tcgaaaaatt   1680
ggtgttttgt ccgcaacgaa ttggaccaaa ttcaaaccaa ctgattattc ggaagaacat   1740
tttacgaaat acaaccgatt cctttctcca atctctacaa tcaagccaca atttaacgaa   1800
tttgatacag tgggtgttgt tgtacaaata actgtgaata gttttgatca aaacttgtgg   1860
ttggcagata attgcaccaa tttgttgttt gttaaaactt tcaatggtcc gaaaagttgt   1920
ctactttttg ataaactcaa agttggacaa gttgttgcaa ttacgaattt aattttattc   1980
```

gataagacaa cagattttgc acaagctgaa gctaatcatt tgacggtcgt ttcatcggca 2040

cccagacaca agtatttgca agaggggctt gatttattgc gtcaacaatt acccaaagaa 2100

ttggacactt ttttgcaaga gtgcaaccag aaagtggctt tttgtaccgt taaaagtgat 2160

ataacgaaaa attacaaaaa ttcgtcaaaa tctctgctac agacgtccga tatttttgat 2220

gataatttgg atagtttcta ctccacgatt gacttcgata aattgaagtg a 2271

<210> 6
<211> 756
<212> PRT
<213> Tribolium castaneum

<400> 6

Met Glu Asp Arg Pro Leu Ser Pro Ile Leu Ser Thr Gln Ile Ser Arg
1 5 10 15

Lys Trp Ile Arg Ser Lys Arg Ser Gln Lys Lys Ser Arg Val Asn Leu
20 25 30

Thr Asn Arg Phe Leu Asp Ala Ser Ser Ser Asn Val Leu Thr Pro Ser
35 40 45

Gln Thr Glu Asn Phe Leu Glu Ser Ile Glu Lys Asp Trp Ser Leu Asp
50 55 60

Lys Asn Ser Asp Asp Val Leu Lys Arg Ala Glu Ile Leu Thr Pro Pro
65 70 75 80

Gln Gly Phe Asn Cys Lys Asn Glu Pro Asn Lys Val Asp Asn Gln Gly
85 90 95

Phe Lys Ala Ser Asp Glu Ala Leu Lys Lys Ala Arg Thr Leu Phe Asp
100 105 110

Asp Met Leu Asn Pro Glu Gly Cys Asn Lys Glu Leu Ser Thr Gly Phe
115 120 125

Ala Gly Phe Lys Thr Ala Ser Gly Lys Asn Ser Leu Arg Ser Glu Asn
130 135 140

Ala Val Lys Glu Ala Arg Asn Arg Asn Ser Asn Glu Gly Thr Thr Thr
145 150 155 160

Ser Phe Val Gly Phe Lys Thr Ala Ser Gly Lys Ser Ser Leu Met Ser
165 170 175

```
Lys Ser Ala Ile Lys Gln Ala Arg Asn Lys Leu Ser Glu Asp Phe Asn
            180                 185                 190

Thr Lys Ile Ser Glu Tyr Asn Ala Ser Phe Ser Gly Phe Lys Thr Ala
        195                 200                 205

Ser Gly Lys Cys Thr Val Met Ser Glu Arg Ala Ile Asn Gln Ala Arg
    210                 215                 220

Asn Lys Leu Asn Glu Asn Phe Asn Phe Glu Asp Val Lys Asn Ala Lys
225                 230                 235                 240

Lys Ser Asp Leu Ser Asp Glu Phe Arg Gly Phe Ser Gln Asp Met Ile
                245                 250                 255

Glu Lys Asn Arg Lys Lys Leu Asp Glu Phe Arg Ser Ile Val Thr Gln
            260                 265                 270

Phe Asn Lys Gly Val Lys Lys Glu Lys Asn Leu Glu Thr Val Arg Glu
        275                 280                 285

Phe Pro Lys Asn Arg Gly Phe Gln Ser Val Asn Leu Ser Lys Asp Arg
    290                 295                 300

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

Arg Lys Arg Arg Phe Phe Glu Gln Glu Lys Phe Gly Phe Ser Pro Val
                325                 330                 335

Arg Lys Asn Pro Ser Val Tyr Asn Ser Thr Pro Leu Lys Asn Thr Ser
            340                 345                 350

Phe Thr Cys Leu Glu Ser Asp Val Thr Pro Ile Lys Lys Ile Lys Asn
        355                 360                 365

Glu Thr Ile Ile Thr Gln Asn Asp Arg Phe Val Ile Gln Ser Cys Thr
    370                 375                 380

Lys Gly Asn Leu Asp Thr Trp Leu Gln Asn Leu Gln Asn Glu Arg Lys
385                 390                 395                 400

Ile Leu Glu Val Lys Leu Lys Val Ile Ile Glu Lys Glu Lys Val Leu
                405                 410                 415
```

```
Asn Leu Gln Lys Glu Ile Leu Glu Asn Arg Val Gln Arg Gln Ser Gly
            420                 425                 430

Ile Leu Phe Thr Arg Lys Glu Asn Ser Glu Arg Val Leu Leu Lys His
        435                 440                 445

Gly Ile Asn Ile Asp Ser Cys Arg Val Ser Ser Glu Ile Asn Ser Glu
    450                 455                 460

Asn Ser Pro Asn Val His Leu Lys Tyr Val Trp Val Trp Glu Lys Leu
465                 470                 475                 480

Pro Met Ile Lys Thr Leu Asp Gly Ala Cys Val Val Pro Asn Val Asp
                485                 490                 495

Asn Phe Val Gly Leu Ser Glu Ile Glu Val Ala Phe Asn Thr Ile Pro
            500                 505                 510

Gly Val Glu Thr Lys Leu Ile Pro Asn Gly Trp Val Gln Asn His Phe
        515                 520                 525

Arg Trp Ile Val Trp Lys Leu Ala Cys Tyr Glu Asn Phe Asn Phe Asp
    530                 535                 540

Cys Leu Thr Val Glu Asn Ile Met Gln Gln Leu Lys Tyr Arg Lys Ile
545                 550                 555                 560

Gly Val Leu Ser Ala Thr Asn Trp Thr Lys Phe Lys Pro Thr Asp Tyr
                565                 570                 575

Ser Glu Glu His Phe Thr Lys Tyr Asn Arg Phe Leu Ser Pro Ile Ser
            580                 585                 590

Thr Ile Lys Pro Gln Phe Asn Glu Phe Asp Thr Val Gly Val Val Val
        595                 600                 605

Gln Ile Thr Val Asn Ser Phe Asp Gln Asn Leu Trp Leu Ala Asp Asn
    610                 615                 620

Cys Thr Asn Leu Leu Phe Val Lys Thr Phe Asn Gly Pro Lys Ser Cys
625                 630                 635                 640

Leu Leu Phe Asp Lys Leu Lys Val Gly Gln Val Val Ala Ile Thr Asn
                645                 650                 655
```

```
Leu Ile Leu Phe Asp Lys Thr Thr Asp Phe Ala Gln Ala Glu Ala Asn
            660                 665                 670

His Leu Thr Val Val Ser Ser Ala Pro Arg His Lys Tyr Leu Gln Glu
        675                 680                 685

Gly Leu Asp Leu Leu Arg Gln Gln Leu Pro Lys Glu Leu Asp Thr Phe
    690                 695                 700

Leu Gln Glu Cys Asn Gln Lys Val Ala Phe Cys Thr Val Lys Ser Asp
705                 710                 715                 720

Ile Thr Lys Asn Tyr Lys Asn Ser Ser Lys Ser Leu Leu Gln Thr Ser
                725                 730                 735

Asp Ile Phe Asp Asp Asn Leu Asp Ser Phe Tyr Ser Thr Ile Asp Phe
            740                 745                 750

Asp Lys Leu Lys
        755
```