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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 April 2004 (29.04.2004)

PCT

(10) International Publication Number
WO 2004/035756 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/033100

(22) International Filing Date: 16 October 2003 (16.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/419,780 17 October 2002 (17.10.2002) US

(71) Applicant *(for all designated States except US)*: **CYTOKINETICS, INC.** [US/US]; 280 East Grand Avenue, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MAXON, Mary** [US/US]; 44 Vicksburg Street, San Francisco, CA 94114 (US). **CHUA, Penelope** [US/US]; 392 San Jose Avenue, San Francisco, CA 94110 (US).

(74) Agents: **STEVENS, Lauren, L.** et al.; Swiss Law Group, Building 3, Palo Alto Square, 3000 El Camino Real, Suite 100, Palo Alto, CA 94306 (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, EG, 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, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, 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, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, 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 2004/035756 A2

(54) Title: MODIFIED NUCLEOTIDE SEQUENCE ENCODING *C. ALBICANS* KiF1 KINESIN PROTEIN

(57) Abstract: The present invention relates to an isolated nucleic acid encoding a CaKif1 kinesin protein wherein the sequence of the isolated nucleic acid has been corrected for optimal expression in an *Escherichia* species or other organism. Disclosed are the nucleic acids, expression vectors, transformants obtained by the transformation with the expression vectors, methods of producing the modified protein using the transformant, and methods of screening for protein modulators.

MODIFIED NUCLEOTIDE SEQUENCE ENCODING *C. ALBICANS* Kif1 KINESIN PROTEIN

CROSS REFERENCE TO RELATED PATENT APPLICATIONS

This application claims priority from U.S. Provisional Patent Application number 60/419,780, filed October 17, 2002, which is incorporated herein by reference for all purposes.

FIELD OF THE INVENTION

[0001] The present invention relates to an isolated nucleic acid encoding a CaKif1 kinesin protein that has been corrected for optimal expression in an *Escherichia* species. Disclosed are the nucleic acids, expression vectors, transformants obtained by the transformation with the expression vector, methods of producing the modified protein using the transformant, and methods of screening for protein modulators.

BACKGROUND OF THE INVENTION

[0002] The kinesin superfamily is an extended family of related microtubule motor proteins. It can be classified into at least 8 subfamilies based on primary amino acid sequence, domain structure, velocity of movement, and cellular function. This family is exemplified by "true" kinesin, which was first isolated from the axoplasm of squid, where it is believed to play a role in anterograde axonal transport of vesicles and organelles (see, e.g., Goldstein, Annu. Rev. Genet. 27:319-351 (1993)). Kinesin uses ATP to generate force and directional movement associated with microtubules.

[0003] Intracellular transport is essential for morphogenesis and functioning of the cell. The kinesin superfamily proteins (KIFs) have been shown to transport membranous organelles and protein complexes in a microtubule- and ATP-dependent manner. More than 30 KIFs have been reported in mice. Almost all kinesins have a region of coiled-coil structure which facilitates dimerization; an exception is the Kif1 family, which is devoid of any substantial coiled-coil region, which together with biophysical data, indicates that these proteins are monomeric (Y. Okada and N.

Hirokawa, Science 283, 1152-1157, 1999). Genetic analysis of mitotic kinesins in yeast and Drosophila, as well as biochemical experiments in other species, have suggested models for the function of these proteins in both mitosis and meiosis.

[0004] *Candida albicans* is one of the most commonly encountered human fungal pathogens, causing a wide variety of infections ranging from mucosal infections in generally healthy persons to life-threatening systemic infections in individuals with impaired immunity. Those at particular risk for such infections are those with AIDS; those having undergone bone marrow or organ transplants; those receiving chemotherapy; and others who have had debilitating illness, severe injury, prolonged hospitalization, or long-term treatment with antibacterial drugs. In order to optimize prevention and treatment of *C. albicans*-related conditions, it is desirable to increase the availability of medically and scientifically important proteins such as CaKif1 for use in clinical therapy and scientific research.

SUMMARY OF THE INVENTION

[0005] The present invention is based on the modification of the sequence encoding the Kif1 kinesin protein in *Candida albicans,* whereby expression of the protein in *E.coli* or other organisms is optimized. The invention provides a codon-corrected *Candida albicans* Kif1 sequence, which when transformed into *E.coli*, provides improved expression.

[0006] In one aspect, the invention provides an isolated nucleic acid sequence encoding a kinesin motor protein, wherein the motor protein (i) has ATPase activity; and (ii) comprises an amino acid sequence of SEQ ID NO: 2, 4, 6, or 8; and wherein the nucleic acid sequence comprises at least one codon that is modified from the *Candida albicans* nucleotide sequence to optimize usage in an *Escherichia* species. In one embodiment, the encoded protein specifically binds to polyclonal antibodies raised against SEQ ID NO: 2, 4, 6, or 8.

[0007] In one embodiment, the nucleic acid has a nucleotide sequence of SEQ ID NO: 1, 3, 5, or 7, wherein the sequence comprises at least one codon that is modified from the *Candida albicans* nucleotide sequence to optimize usage in an *Escherichia* species, and the nucleic acid also encodes a protein having ATPase activity.

[0008] In another aspect, the invention provides an isolated protein comprising an amino acid sequence of SEQ ID NO: 2, 4, 6, or 8, wherein the protein has ATPase activity.

[0009] In another aspect, the invention provides an expression vector, wherein the vector has one or more of the properties described above.

[0010] In a further aspect, the present invention provides a transformed host cell transformed with the expression vector aforementioned. In one embodiment, the transformed host cell is *Escherichia coli.*

[0011] In one embodiment, the present invention provides a method for producing the modified kinesin protein, comprising the steps of:

providing an expression vector,

transforming compatible *E. coli* host cells with the vector, thereby obtaining a transformant;

culturing the transformed cells under conditions suitable for growth; and

recovering the protein from the host cell culture.

[0012] In another embodiment, the present invention provides a method of identifying a compound as a modulator of an activity of the protein. The method comprises contacting the protein with a compound at a first concentration and determining a level of activity of the target protein. The method further comprises contacting the protein with a compound at a second concentration, and then determining a level of activity of the protein. A change in the level of activity between the protein contacted with the first concentration and the second concentration indicates that the compound modulates an activity of the protein.

[0013] Also provided are modulators of the protein including agents for the treatment of fungal infection as common in conditions of immunocompromise or impaired immunity, including AIDS, bone marrow or organ transplants, chemotherapy, or prolonged antibacterial therapy. The agents and compositions provided herein can be used in a variety of applications and formulations.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

[0014] The Sequence Listing, which is incorporated herein by reference in its entirely, provides exemplary sequences including polynucleotide sequences SEQ ID NOs: 1, 3, 5, and 7, and polypeptide sequences, SEQ ID NOs: 2, 4, 6, and 8.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0015] "Allele" refers to any of two or more alternative forms of a gene occupying the same chromosomal locus.

[0016] "ADP" refers to adenosine diphosphate and also includes ADP analogs, including, but not limited to, deoxyadenosine diphosphate (dADP) and adenosine analogs.

[0017] "Antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively. The term antibody also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant DNA methodologies.

[0018] "Variant" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCT all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the

corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each degenerate codon in a nucleic acid can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

[0019] Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to about 20 amino acids, although considerably longer insertions may be tolerated. Deletions range from about 1 to about 20 residues, although in some cases, deletions may be much longer.

[0020] Substitutions, deletions, and insertions or any combinations thereof may be used to arrive at a final derivative. Generally, these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger characteristics may be tolerated in certain circumstances.

[0021] The following six groups each contain amino acids that are conservative substitutions for one another:

Alanine (A), Serine (S), Threonine (T);

Aspartic acid (D), Glutamic acid (E);

Asparagine (N), Glutamine (Q);

Arginine (R), Lysine (K);

Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and

Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(see, e.g., Creighton, Proteins (1984)).

[0022] "Cytoskeletal component" denotes any molecule that is found in association with the cellular cytoskeleton, that plays a role in maintaining or regulating the structural integrity of the cytoskeleton, or that mediates or regulates motile events mediated by the cytoskeleton. Includes cytoskeletal polymers (e.g., actin filaments, microtubules, intermediate filaments, myosin fragments), molecular motors (e.g., kinesins, myosins, dyneins), cytoskeleton associated regulatory proteins (e.g.,

tropomysin, alpha-actinin) and cytoskeletal associated binding proteins (e.g., microtubules associated proteins, actin binding proteins).

[0023] "Cytoskeletal function" refers to biological roles of the cytoskeleton, including but not limited to the providing of structural organization (e.g., microvilli, mitotic spindle) and the mediation of motile events within the cell (e.g., muscle contraction, mitotic chromosome movements, contractile ring formation and function, pseudopodal movement, active cell surface deformations, vesicle formation and translocation.)

[0024] A "diagnostic" as used herein is a compound, method, system, or device that assists in the identification and characterization of a health or disease state. The diagnostic can be used in standard assays as is known in the art.

[0025] An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

[0026] "High stringency conditions" may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.2 times SSC and 0.1% sodium dodecyl sulfate at 68°C.; (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C.; or (3) employ 50% formamide, 5.times.SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5.times.Denhardt's solution, sonicated salmon sperm DNA (50 .mu.g/ml), 0.1% SDS, and 10% dextran sulfate at 42.degree. C., with washes in 0.2.times.SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1.times.SSC containing EDTA at 55°C.

[0027] By "host cell" is meant a cell that contains an expression vector and supports the replication or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or

mammalian cells such as CHO, HeLa and the like, or plant cells. Both primary cells and cultured cell lines are included in this definition.

[0028] The terms "isolated", "purified", or "biologically pure" refer to material that is substantially or essentially free from components which normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In an isolated gene, the nucleic acid of interest is separated from open reading frames which flank the gene of interest and encode proteins other than the protein of interest. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

[0029] "Modulators," "inhibitors," and "activators of a target protein" refer to modulatory molecules identified using in vitro and in vivo assays for target protein activity. Such assays include ATPase activity, microtubule gliding, microtubule depolymerizing activity, and binding activity such as microtubule binding activity or binding of nucleotide analogs. Samples or assays that are treated with a candidate agent at a test and control concentration. The control concentration can be zero. If there is a change in target protein activity between the two concentrations, this change indicates the identification of a modulator. A change in activity, which can be an increase or decrease, is preferably a change of at least 20% to 50%, more preferably by at least 50% to 75%, more preferably at least 75% to 100%, and more preferably 150% to 200%, and most preferably is a change of at least 2 to 10 fold compared to a control. Additionally, a change can be indicated by a change in binding specificity or substrate.

[0030] The phrase "motor domain" refers to the domain of a target protein that confers membership in the kinesin superfamily of motor proteins through a sequence identity of approximately 35-45% identity to the motor domain of true kinesin.

[0031] The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural

nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. For example, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260)2605-2608 (1985); Cassol et al. 1992; Rossolini et al. Mol. Cell. Probes 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

[0032] "Nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e., A, G, C, or T) or modified bases. In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidine complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

[0033] The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residues is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. A target protein comprises a polypeptide demonstrated to have at least microtubule stimulated ATPase activity. Amino acids may be referred to herein by either

their commonly known three letter symbols or by Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes, i.e., the one-letter symbols recommended by the IUPAC-IUB.

[0034] A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA box element. A promoter also optionally includes distal enhancer or repressor elements which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The Target Protein

[0035] The present invention provides for the modification of the nucleic acid sequence encoding a CaKif1 kinesin protein, corrected for optimal expression in an *Escherichia* species or other organism. The corrected codon is the CUG codon. *Candida albicans* differs from most yeast strains in that it does not use the same genetic code. The codon CUG, which in the universal code is read as leucine, is decoded as a serine in *Candida.* Replacement of leucine by serine at CUG encoded residues has been shown to often result in the production of inactive proteins.

[0036] In one aspect, CaKif1 can be defined by having at least one or preferably more than one of the following functional and structural characteristics. Functionally, CaKif1 will have microtubule-stimulated ATPase activity, and microtubule motor activity that is ATP dependent. CaKif1 activity can also be described in terms of its ability to bind microtubules.

[0037] The modification of the nucleotide sequences provided herein encode CaKif1 or fragments thereof. Thus, in one aspect, the nucleic acids provided herein are

defined by the proteins herein. The protein provided herein comprises an amino acid sequence which has one or more of the following characteristics: greater than 70% sequence identity with SEQ ID NO:2 or SEQ ID NO:4, preferably greater than 80%, more preferably greater than 90%, more preferably greater than 95% or, in another embodiment, has 98 to 100% sequence identity with SEQ ID NO:2 or SEQ ID NO:4. As described above when describing the nucleotide in terms of SEQ ID NO:1, the sequence identity may be slightly lower due to the degeneracy in the genetic code. Also included within the definition of the target proteins are amino acid sequence variants of wild-type target proteins.

[0038] Portions of the modified CaKif1 nucleotide sequence may be used to identify polymorphic variants, orthologs, alleles, and homologues of CaKif1. This identification can be made in vitro, e.g., under stringent hybridization conditions and sequencing, or by using the sequence information in a computer system for comparison with other nucleotide sequences.

[0039] As will be appreciated by those in the art, the target proteins can be made in a variety of ways, including both synthesis de novo and by expressing a nucleic acid encoding the protein.

[0040] Target proteins of the present invention may also be modified in a way to form chimeric molecules comprising a fusion of a target protein with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino or carboxyl terminus of the target protein. Provision of the epitope tag enables the target protein to be readily detected, as well as readily purified by affinity purification. Various tag epitopes are well known in the art. Examples include polyhistidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 (see, Field et al. (1988) Mol. Cell. Biol.:2159); the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto (see, Evans et al., (1985) Molecular and Cellular Biology, 5:3610); and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody (see, Paborsky et al., (1990) Protein Engineering 3:547). Other tag polypeptides include the Flag-peptide (see, Hopp, et al. (1988) BioTechnology 6:1204); the Kt3 epitope peptide (see, Martine et al. (1992) Science, 255:192); tubulin epitope peptide (see, Skinner (1991) J. Biol. Chem.

266:15173); and the T7 gene 10 protein peptide tag (see, Lutz-Freyermuth et al. (1990) Proc. Natl. Acad. Sci. USA 87:6393).

[0041] The biological activity of any of the peptides provided herein can be routinely confirmed by assays such as those which assay ATPase activity or microtubule binding activity. In one embodiment, polymorphic variants, alleles, and orthologs, homologues of the modified CaKif1 are confirmed by using ATPase or microtubule binding assays as known in the art.

[0042] The isolation of the biologically active modified CaKif1 provides a means for assaying for modulators of this kinesin superfamily protein. Biologically active modified CaKif1 is useful for identifying modulators of CaKif1 or fragments thereof and kinesin superfamily members using in vitro assays such as microtubule gliding assays, ATPase assay (Kodama et al., J. Biochem. 99;1465-1472 (1986); Stewart et al., Proc. Nat'l Acad. Sci USA 90: 5209-5213 (1993)), and binding assays including microtubule binding assays (Vale et al., Cell 42: 39-50 (1985)).

Isolation of the Gene Encoding CaKif1

General Recombinant DNA Methods

[0043] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook et al., Molecular Cloning, A Laboratory Manual (2nd ed. 1989); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 1994).

[0044] For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa) or amino acid residue numbers. Protein sizes are estimated from gel electrophoresis, mass spectrometry, sequenced proteins, derived amino acid sequences, or from published protein sequences.

[0045] Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described

by Beaucage & Caruthers, Tetrahedron Letts. 22:1859-1962 (1981), using an automated synthesizer, as described in Van Devanter et al., Nucleic Acids Res. 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson & Reanier, J. Chrom. 225:137-149 (1983).

[0046] The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace et al., Gene 16:21-26 (1981).

Cloning Methods for the Isolation of Modified Nucleotide Sequences Encoding Kif1 Protein in *Candida Albicans*.

[0047] In general, the modified nucleic acid sequences encoding CaKif1 and related nucleic acid sequence homologs are cloned from cDNA and genomic DNA libraries or isolated using amplification techniques with oligonucleotide primers. Alternatively, expression libraries can be used to clone the modified CaKif1 and CaKif1 homologues by detected expressed homologues immunologically with antisera or purified antibodies made against CaKif1 that also recognize and selectively bind to the CaKif1 homologue. Finally, amplification techniques using primers can be used to amplify and isolate CaKif1 from DNA or RNA. Amplification techniques using degenerate primers can also be used to amplify and isolate CaKif1 homologues. Amplification techniques using primers can also be used to isolate a nucleic acid encoding CaKif1. These primers can be used, e.g., to amplify a probe of several hundred nucleotides, which is then used to screen a library for full-length CaKif1.

[0048] Appropriate primers and probes for identifying the gene encoding homologues of CaKif1 in other species are generated from comparisons of the sequences provided herein. As described above, antibodies can be used to identify CaKif1 homologues. For example, antibodies made to the motor domain of CaKif1 or to the whole protein are useful for identifying CaKif1 homologues.

[0049] To make a cDNA library, one should choose a tissue source that is rich in the mRNA of choice. For a genomic library, the DNA is extracted from the tissue and

either mechanically sheared or enzymatically digested to yield fragments of about 12-20 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged in vitro. Recombinant phage are analyzed by plaque hybridization as described in Benton & Davis, Science 196:180-182 (1977). Colony hybridization is generally described in Grunstein et al., Proc. Natl. Acad. Sci. USA, 72:3961-3965 (1975).

[0050] An alternative method of isolating CaKif1 modified nucleic acid and its homologues combines the use of synthetic oligonucleotide primers and amplification of an RNA or DNA template (see U.S. Pat. Nos. 4,683,195 and 4,683,202; PCR Protocols: A Guide to Methods and Applications (Innis et al., eds. 1990)). Methods such as polymerase chain reaction and ligase chain reaction can be used to amplify nucleic acid sequences of CaKif1 directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. Degenerate oligonucleotides can be designed to amplify CaKif1 homologues using the sequences provided herein. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other in vitro amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of CaKif1 encoding mRNA in physiological samples, for nucleic sequencing or for other purposes. Genes amplified by the PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

[0051] Gene expression of CaKif1 can also be analyzed by techniques known in the art, e.g., reverse transcription and amplification of mRNA, isolation of total RNA or poly A+RNA, northern blotting, dot blotting, *in situ* hybridization, RNase protection, quantitative PCR, and the like.

[0052] Synthetic oligonucleotides can be used to construct recombinant CaKif1 genes for use as probes or for expression of protein. This method is performed using a series of overlapping oligonucleotides usually 40-120 bp in length, representing both the sense and nonsense strands of the gene. These DNA fragments are then annealed, ligated and cloned. Alternatively, amplification techniques can be used with precise primers to amplify a specific subsequence of the CaKif1 gene. The specific subsequence is then ligated into an expression vector.

[0053] The gene for the modified CaKif1 protein is typically cloned into intermediate vectors before transformation into prokaryotic or eukaryotic cells for replication and/or expression. The intermediate vectors are typically prokaryote vectors or shuttle vectors.

Expression Vector in Prokaryotic Host Cell

[0054] To obtain high level expression of a cloned gene, such as those cDNAs encoding the modified sequence of the CaKif1 protein, it is important to construct an expression vector that contains a strong promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and described, e.g., in Sambrook et al. and Ausubel et al. Bacterial expression systems for expressing the CaKif1 protein are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., Gene 22:229-235 (1983); Mosbach et al., Nature 302:543-545 (1983)). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available. The pET expression system (Novagen) is a preferred prokaryotic expression system.

[0055] The promoter used to direct expression of a heterologous nucleic acid depends on the particular application. The promoter is preferably positioned about the same distance from the heterologous transcription start site as it is from the transcription start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

[0056] In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the CaKif1 encoding the modified nucleic acid in host cells. A typical expression cassette thus contains a promoter operably linked to the nucleic acid sequence encoding modified CaKif1 and signals required for efficient polyadenylation of the transcript, ribosome binding sites, and translation termination. The modified nucleic acid sequence encoding CaKif1 may typically be linked to a cleavable signal peptide sequence to promote secretion of the encoded protein by the transformed cell. Such signal

peptides would include, among others, the signal peptides from tissue plasminogen activator, insulin, and neuron growth factor, and juvenile hormone esterase of *Heliothis virescens*. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

[0057] In addition to a promoter sequence, the expression cassette should also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from different genes.

[0058] The particular expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic or prokaryotic cells may be used. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23, and fusion expression systems such as GST and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, e.g., c-myc or histidine tags.

[0059] The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of eukaryotic sequences. The particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable. The prokaryotic sequences are preferably chosen such that they do not interfere with the replication of the DNA in eukaryotic cells, if necessary.

Transformation Methods

[0060] Standard transfection or transformation methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of the modified CaKif1 protein, which are then purified using standard techniques (see, e.g., Colley et al., J. Biol. Chem. 264:17619-17622 (1989); Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher ed., 1990)).

[0061] Transformation of the prokaryotic *E. coli* host cell is performed according to standard techniques (see, e.g., Morrison, J. Bact., 132:349-351 (1977); Clark-Curtiss & Curtiss, Methods in Enzymology, 101:347-362 (Wu et al., eds, 1983). Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, plasma vectors, viral Vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the modified CaKif1.

[0062] The media for culturing the transformants are well-known. For culturing *E. coli*, a nutrient medium such as LB medium or a minimal medium such as M9 medium is used with the addition of a carbon source, a nitrogen source, a vitamin source, etc. The transformant is cultured at approximately 16 to 42 degrees C. for 5-168 hours. The culture conditions may vary. *E. coli* cultures will be aerated on a shaker.

[0063] After the expression vector is introduced into the cells, the transfected cells are cultured under conditions favoring expression of the modified CaKif1, which is recovered from the culture using standard techniques.

Host Bacterial Strains

[0064] Numerous strains of *E. coli* exist which may serve as a host. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31, 446); *E. coli* X1776 (ATCC 31,537); *E. coli* Strain W3110 (ATCC 27,325); and K5 772 (ATCC 53,635). Preferred bacterial strains include *E. coli* BL21 (DE3), BL21 9DE3), pLysS, BL21 (DE3) pLysEIV. Also used are derivatives of BL21 (DE3) codon plus (Stratagene), Rosetta (Novagen), and star strains (Stratagene). Codon plus and Rosetta express rare codons, resulting in better expression of human proteins in *E. coli*. The star strains have an RNAse gene deleted for higher mRNA stability and therefore, higher protein expression.

Purification of the Modified CaKif1 Protein

[0065] The modified CaKif1 protein may be purified for use in functional assays. In a preferred embodiment, the protein is purified for use in assays to provide substantially pure samples. Alternatively, the protein need not be substantially pure as long as the sample comprising the target protein is substantially free of other components that can contribute to the production of ADP or phosphate.

[0066] The protein may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample. Standard purification methods include electrophoretic, molecular, immunological, and chromatographic techniques, including ion exchange, hydrophobic, affinity, and reverse-phase HPLC chromatography, chromatofocussing, selective precipitation with such substances as ammonium sulfate; and others (see, e.g., Scopes, Protein Purification: Principles and Practice (1982); U.S. Pat. No. 4,673,641; Ausubel et al. supra; and Sambrook et al., supra). For example, the target protein can be purified using a standard anti-target antibody column. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. A preferred method of purification is use of Ni-NTA agarose (Qiagen).

[0067] The expressed protein can be purified by standard chromatographic procedures to yield a purified, biochemically active protein. The activity of any of the peptides provided herein can be routinely confirmed by the assays provided herein such as those which assay ATPase activity or microtubule binding activity. Biologically active target protein is useful for identifying modulators of target protein or fragments thereof and kinesin superfamily members using in vitro assays such as microtubule gliding assays, ATPase assays (Kodama et al., J. Biochem. 99:1465-1472 (1986); Stewart et al., Proc. Nat'l Acad. Sci. USA 90:5209-5213 (1993)), and binding assays including microtubule binding assays (Vale et al., Cell 42:39-50 (1985)), as described in detail below.

Purification of the Modified CaKif1 Protein from Recombinant Bacteria

[0068] Recombinant proteins are expressed by transformed bacteria in large amounts, typically after promoter induction; but expression can be constitutive. Promoter

induction with IPTG is a preferred method of expression. Bacteria are grown according to standard procedures in the art. Fresh or frozen bacteria cells are used for isolation of protein. Alternatively, it is possible to purify CaKif1 from bacteria periplasm. After CaKif1 is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to skill in the art. To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM MgSO.sub.4 and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

[0069] Suitable purification schemes for some specific kinesins are outlined in U.S. Ser. No. 09/295,612, filed Apr. 20, 1999, hereby expressly incorporated herein in its entirety for all purposes.

Standard Protein Separation Techniques for Purifying Modified CaKif1 Protein

Solubility Fractionation

[0070] Often as an initial step, particularly if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol includes adding saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This concentration will precipitate the most hydrophobic of proteins. The precipitate is then discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer

and the excess salt removed if necessary, either through dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

Size Differential Filtration

[0071] The molecular weight of the modified CaKif1 protein can be used to isolate it from proteins of greater and lesser size using ultrafiltration through membranes of different pore size (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

Column Chromatography

[0072] The modified CaKif1 protein can also be separated from other proteins on the basis of its size, net surface charge, hydrophobicity, and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art. It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., Pharmacia Biotech).

Immunological Detection of Modified CaKif1 Protein

[0073] In addition to the detection of the modified CaKif1 genes and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect modified CaKif1. Immunoassays can be used to qualitatively or quantitatively analyze CaKif1. A general overview of the applicable technology can be found in Harlow & Lane, Antibodies: A Laboratory Manual (1988).

Antibodies to Modified CaKif1

[0074] Methods of producing polyclonal and monoclonal antibodies that react specifically with CaKif1 are known to those of skill in the art (see, e.g., Coligan, Current Protocols in Immunology (1991); Harlow & Lane, supra; Goding, Monoclonal Antibodies: Principles and Practice (2d ed. 1986); and Kohler & Milstein, Nature 256:495-497 (1975). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (see, e.g., Huse et al., Science 246:1275-1281 (1989); Ward et al., Nature 341:544-546(1989)).

[0075] A number of CaKif1 protein comprising immunogens may be used to produce antibodies specifically reactive with CaKif1. For example, recombinant CaKif1 or an antigenic fragment thereof such as the motor domain, is isolated as described herein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells as described above, and purified as generally described above. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Naturally occurring protein may also be used either in pure or impure form. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated, for subsequent use in immunoassays to measure the protein.

[0076] Methods of production of polyclonal antibodies are known to those of skill in the art. An inbred strain of mice (e.g., BALB/C mice) or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to CaKif1. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (see Harlow & Lane, supra).

[0077] Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see Kohler & Milstein, Eur. J. Immunol. 6:511-519 (1976)). Alternative methods of immortalization

include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al., Science 246:1275-1281 (1989).

[0078] Monoclonal antibodies and polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Typically, polyclonal antisera with a titer of 10.sup.4 or greater are selected and tested for their cross reactivity against non-CaKif1 proteins or even other homologous proteins from other organisms (e.g., *C. elegans* unc-104 or human Kif1A), using a competitive binding immunoassay. Specific polyclonal antisera and monoclonal antibodies will usually bind with a K.sub.d of at least about 0.1 mM, more usually at least about 1 .mu.M, preferably at least about 0.1 .mu.M or better, and most preferably, 0.01 .mu.M or better.

[0079] Once modified CaKif1 protein specific antibodies are available, CaKif1 can be detected by a variety of immunoassay methods. For a review of immunological and immunoassay procedures, see Basic and Clinical Immunology (Stites & Terr eds., 7th ed. 1991). Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Enzyme Immunoassay (Maggio ed., 1980); and Harlow & Lane, supra.

Binding Assays

[0080] Antibodies can be used for treatment or to identify the presence of modified CaKif1 protein having the sequence identity characteristics as described herein. Additionally, antibodies can be used to identify modulators of the interaction between the antibody and CaKif1 as further described below. While the following discussion is directed toward the use of antibodies in the use of binding assays, it is understood that the same general assay formats such as those described for "non-competitive" or

"competitive" assays can be used with any compound which binds to the modified CaKif1 such as microtubules or the compounds described in Ser. No. 60/070,772.

[0081] In a preferred embodiment, the modified CaKif1 protein is detected and/or quantified using any of a number of well recognized immunological binding assays (see, e.g., U.S. Pat. Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also Methods in Cell Biology Volume 37: Antibodies in Cell Biology (Asai, ed. 1993); Basic and Clinical Immunology (Stites & Terr, eds., 7th ed. 1991): Immunological binding assays (or immunoassays) typically use an antibody that specifically binds to a protein or antigen of choice (in this case the CaKif1 or antigenic subsequence thereof). The antibody (e.g., anti-CaKif1) may be produced by any of a number of means well known to those of skill in the art and as described above.

Other Assay Formats

[0082] Western blot (immunoblot) analysis is used to detect and quantify the presence of modified CaKif1 protein in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind the modified CaKif1. The anti-CaKif1 antibodies specifically bind to the CaKif1 on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (e.g., labeled sheep anti-mouse antibodies) that specifically bind to the anti-CaKif1 antibodies. Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (e.g., antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (see Monroe et al., Amer. Clin. Prod. Rev. 5:34-41 (1986)).

Reduction of Non-specific Binding

[0083] One of skill in the art will appreciate that it is often desirable to minimize non-specific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of

non-specific binding to the substrate. Means of reducing such non-specific binding are well known to those of skill in the art. Typically, this technique involves coating the substrate with a proteinaceous composition. In particular, protein compositions such as bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used with powdered milk being most preferred.

Assays for Modulators of the Target Protein

Functional Assays

[0084] Assays that can be used to test for modulators of the target protein include a variety of *in vitro* or *in vivo* assays, e.g., microtubule gliding assays, binding assays such as microtubule binding assays, microtubule depolymerization assays, and ATPase assays (Kodama et al., J. Biochem. 99: 1465-1472 (1986); Stewart et al., Proc. Nat'l Acad. Sci. USA 90: 5209-5213 (1993); (Lombillo et al., J. Cell Biol. 128:107-115 (1995); (Vale et al., Cell 42:39-50 (1985)).

[0085] Modulation is tested by screening for candidate agents capable of modulating the activity of the target protein comprising the steps of combining a candidate agent with the target protein, as above, and determining an alteration in the biological activity of the target protein. Thus, in this embodiment, the candidate agent should both bind to the target protein (although this may not be necessary), and alter its biological or biochemical activity as defined herein. The methods include both in vitro screening methods and in vivo screening of cells for alterations in cell cycle distribution, cell viability, or for the presence, morphology, activity, distribution, or amount of mitotic spindles, as are generally outlined above.

[0086] In a preferred embodiment, molecular motor activity is measured by the methods disclosed in U. S. Patent 6, 410,254 (Finer, J., et al..), entitled "Compositions and assay utilizing ADP or phosphate for detecting protein modulators," which is incorporated herein by reference in its entirety. More specifically, this assay detects modulators of any aspect of a kinesin motor function ranging from interaction with microtubules to hydrolysis of ATP. In one embodiment, ADP or phosphate is used as the readout for protein activity.

[0087] There are a number of enzymatic assays known in the art which use ADP as a substrate. For example, kinase reactions such as pyruvate kinases are known. See, Nature 78:632 (1956) and Mol. Pharmacol. 6:31 (1970). This is a preferred method in that it allows the regeneration of ATP. In one embodiment, the level of activity of the enzymatic reaction is determined directly. In a preferred embodiment, the level of activity of the enzymatic reaction which uses ADP as a substrate is measured indirectly by being coupled to another reaction. For example, in one embodiment, the method further comprises a lactate dehydrogenase reaction under conditions which normally allow the oxidation of NADH, wherein said lactate dehydrogenase reaction is dependent on the pyruvate kinase reaction. Measurement of enzymatic reactions by coupling is known in the art. Furthermore, there are a number of reactions which utilize phosphate. Examples of such reactions include a purine nucleoside phosphorylase reaction. This reaction can be measured directly or indirectly. A particularly preferred embodiments utilizes the pyruvate kinase/lactate dehydrogenase system.

[0088] In one embodiment, the detection of the ADP or phosphate proceeds non-enzymatically, for example, by binding or reacting the ADP or phosphate with a detectable compound. For example, phosphomolybdate based assays may be used which involve conversion of free phosphate to a phosphomolybdate complex. One method of quantifying the phosphomolybdate is with malachite green. Alternatively, a fluorescently labeled form of a phosphate binding protein, such as the *E. coli* phosphate binding protein, can be used to measure phosphate by a shift in its fluorescence.

[0089] In addition, target protein activity can be examined by determining modulation of target protein in vitro using cultured cells. The cells are treated with a candidate agent and the effect of such agent on the cells is then determined whether directly or by examining relevant surrogate markers. For example, characteristics such as mitotic spindle morphology and cell cycle distribution can be used to determine the effect.

[0090] Thus, in a preferred embodiment, the methods comprise combining a target protein and a candidate, and determining the effect of the candidate on the target protein. Generally, a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically,

one of these concentrations serves as a negative control, i.e., at zero concentration or below the level of detection.

[0091] As will be appreciated by those in the art, the components may be added in buffers and reagents to assay target protein activity and give optimal signals. Since the methods allow kinetic measurements, the incubation periods can be optimized to give adequate detection signals over the background.

[0092] In a preferred embodiment, an antifoam or a surfactant is included in the assay mixture. Suitable antifoams include, but are not limited to, antifoam 289 (Sigma). Suitable surfactants include, but are not limited to, Tween, Tritons, including Triton X-100, saponins, and polyoxyethylene ethers. Generally, the antifoams, detergents, or surfactants are added at a range from about 0.01 ppm to about 10 ppm.

Binding Assays

[0093] Competitive screening assays may be done by combining the target protein and a compound in a first sample in a first concentration. A level of activity is then determined for the protein. The protein is further contacted with the compound at a second concentration, and a level of activity of the protein is determined. A difference between the level of activity of the protein contacted with the first concentration of the compound and the level of activity of the protein contacted with the second concentration of the compound indicates that the compound is a modulator of protein activity.

Other Assay Components

[0094] The assays provided utilize target protein as defined herein. In one embodiment, portions of target protein are utilized; in a preferred embodiment, portions having target protein activity as described herein are used. In addition, the assays described herein may utilize either isolated target proteins or cells or animal models comprising the target proteins.

[0095] A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also, reagents that otherwise improve the efficiency of the

assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc.; may be used. The mixture of components may be added in any order that provides for the requisite binding.

Cell-Based Assays

[0096] A variety of cell-based assays may be used to determine activity. Among these are growth inhibition, halo assays, disc plate diffusion, and inhibition of fungal hyphae length. These assays utilize standard techniques that are well-known in the art ((R.N. Jones et al, Manual of Clinical Microbiology, 4th ed., (1985); and M.A. Pfaller et al, Antimicrobial Agents and Chemotherapy, 34 (1990)).

Applications

[0097] The methods of the invention are used to identify compounds useful in the treatment of systemic fungal infections in mammals.

[0098] Candidate agents having the desired pharmacological activity may be administered in a physiologically acceptable carrier to a patient, as described herein. Depending upon the manner of introduction, the compounds may be formulated in a variety of ways as discussed below. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt. %. The agents maybe administered alone or in combination with other treatments, i.e., radiation, or other chemotherapeutic agents.

[0099] In a preferred embodiment, the pharmaceutical compositions are in a water soluble form, such as pharmaceutically acceptable salts, which is meant to include both acid and base addition salts.

[0100] The pharmaceutical compositions can be prepared in various forms, such as granules, tablets, pills, suppositories, capsules, suspensions, salves, lotions and the like. Pharmaceutical grade organic or inorganic carriers and/or diluents suitable for oral and topical use can be used to make up compositions containing the therapeutically-active compounds. Diluents known to the art include aqueous media, vegetable and animal oils and fats. Stabilizing agents, wetting and emulsifying agents, salts for varying

the osmotic pressure or buffers for securing an adequate pH value, and skin penetration enhancers can be used as auxiliary agents. The pharmaceutical compositions may also include one or more of the following: carrier proteins such as serum albumin; buffers; fillers such as microcrystalline cellulose, lactose, corn and other starches; binding agents; sweeteners and other flavoring agents; coloring agents; and polyethylene glycol. Additives are well known in the art, and are used in a variety of formulations.

[0101] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety.

CLAIMS

What is claimed is:

1. An isolated nucleic acid encoding a kinesin motor protein, wherein (i) the protein has ATPase activity; and (ii) the protein comprises an amino acid sequence of SEQ ID NO: 2, 4, 6, or 8; and wherein the nucleic acid comprises at least one codon that is modified from the *Candida albicans* nucleotide sequence to optimize usage in an *Escherichia* species.

2. The isolated nucleic acid of claim 1, wherein the encoded protein specifically binds to polyclonal antibodies generated against SEQ ID NO: 2, 4, 6, or 8.

3. The isolated nucleic acid of claim 1, wherein said nucleic acid encodes the protein having the sequence of SEQ ID NO: 2.

4. The isolated nucleic acid of claim 1, wherein said nucleic acid encodes the protein having the sequence of SEQ ID NO: 4.

5. The isolated nucleic acid of claim 1, wherein said nucleic acid encodes the protein having the sequence of SEQ ID NO: 6.

6. The isolated nucleic acid of claim 1, wherein said nucleic acid encodes the protein having the sequence of SEQ ID NO: 8.

7. An isolated nucleic acid comprising a nucleotide sequence of SEQ ID NO:1, 3, 5, or 7, wherein the nucleic acid sequence comprises at least one codon that is modified from the *Candida albicans* nucleotide sequence to optimize usage in an *Escherichia* species and wherein the nucleic acid encodes a protein having ATPase activity.

8. The isolated nucleic acid of claim 7, wherein said nucleic acid has the sequence of SEQ ID NO: 1.

9. The isolated nucleic acid of claim 7, wherein said nucleic acid hs the sequence of SEQ ID NO: 3.

10. The isolated nucleic acid of claim 7, wherein said nucleic acid has the sequence of SEQ ID NO: 5.

11. The isolated nucleic acid of claim 7, wherein said nucleic acid has the sequence of SEQ ID NO: 7.

12. An isolated protein comprising an amino acid sequence of SEQ ID NO: 2, 4, 6, or 8, wherein the protein has ATPase activity.

13. The isolated protein of claim 12, wherein said protein comprises the amino acid sequence of SEQ ID NO: 2.

14. The isolated protein of claim 12, wherein said protein comprises the amino acid sequence of SEQ ID NO: 4.

15. The isolated protein of claim 12, wherein said protein comprises the amino acid sequence of SEQ ID NO: 6.

16. The isolated protein of claim 12, wherein said protein comprises the amino acid sequence of SEQ ID NO: 8.

17. An expression vector comprising an isolated nucleic acid of claim 1 or 7.

18. A host cell transfected with the vector of claim 17.

19. The transformed host of claim 18, which is *Escherichia coli*.

20. A method for producing the protein of claim 12, the method comprising the steps of:

culturing the host cell of claim 18 under conditions suitable for the expression of the protein; and

recovering the protein from the host cell culture.

21. A method of identifying a compound that modulates an activity of a protein of claim 12, the method comprising:

contacting the protein with a compound at a first concentration and

determining a level of activity of the protein; and

contacting the protein with said compound at a second concentration, and

determining a level of activity of the protein,

wherein a difference between the level of activity of the protein contacted with the first concentration of the compound and the level of activity of the protein contacted with the second concentration of the compound indicates that the compound modulates the activity of the protein.

22. The method of Claim 21 wherein said activity of said protein is determined by an ATPase assay, a microtubule gliding assay, a binding assay or a microtubule depolymerization assay."

SEQUENCE LISTING

<110> Cytokinetics, Inc.
Maxon, Mary
Chua, Penelope

<120> MODIFIED NUCLEOTIDE SEQUENCE ENCODING C. ALBICANS Kif1 KINESIN PROTEIN

<130> CYTOP075WO

<150> 60/419,780
<151> 2002-10-17

<160> 8

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1617
<212> DNA
<213> Candida Albicans

<400> 1
atgtcaaaca atatcaaagt atcagttcga gttagacctt tattgccacg ggagctaaaa 60
gctaaaaagg ccgaagggga ggccacacca gtatcattag tgtcaatgcc cccgtcagat 120
cctcataaag ttctattgac caaccccaag gcacaatcga agaacaatat cagtattgac 180
aacaagccca agagtttttt atttgatgag tgtatttggt catttaatag taaggattcg 240
aattacactg ataatcacaa atattatgaa aaaaccggac cacaattgtt gtcccatttc 300
tttgaagggt ataatgtttg ccttttggcg tatggtcaaa caggatcagg gaaaacatat 360
acaatgatgg gtgataaaaa tgaatcgcca gggataatac cattgttgat aaaagatatc 420
ttaaagcaaa tggagatttc tgttaatgag aaaatcaatt gtgaattgaa gttttcgtac 480
attgaaatat acaatgagca agtcaaagat ttgcttacag gtgacaccaa aaatgggaac 540
gggaatggga acagtacaca tgcaagccaa aagtgcaaag ttagagaaca tcctgttact 600
ggtccatacg tcgagaatgt gaaagaatac aatatcgaaa attataccca atttttgaaa 660
ttgttggcta aaggtaattc tagtcgttca acagcatcta catctatgaa cgataaaagt 720
tcaagaagtc acgcaattat cacattgaca ttaaagcaaa ccaaatttga aaaatttagc 780
ggttttagca gtgattctga tgaaaccaac agtattggtg atgcaagtga agagatggta 840
tcgaacatca agttggtcga tcttgctgga tcagaaagac tacagaaaac aaaattatat 900
ggacaacaag aaagaatcaa agaagggaca ttgattaaca aatcattggc tgttttggga 960
agatgtataa atttgctttc aagcaatcca tcgtcattga tcccctaccg tgaatcaatt 1020
ttaacatatt tattaaaaga aaatcttggt gggaattcta aaacatgcat gatcttttgt 1080
gtttctccag ttgactatga agaaacccat caaacattga actacgccaa tgaagtgaaa 1140
aaaatcaaaa cactagccaa ggcaaataaa acaaaacttt caaatgttcc tattgattgg 1200
caacagttac aatcgacaga taaaacagtt atccaatctt tgaagaatga aattgacgtt 1260
ttaacaaata aattaaatca aatgcaacaa ttaaatgaac ctagatctca tccaccacaa 1320
tcaataaata atataattga atatttggat aaagaaacta ataaagtgaa atttgaaaat 1380
aaatatttga aacaaatgat gaataccaaa aacaatcaca taaaaggatt gaacaatcac 1440
attgatttta tttcgaacga atatcaattg ttgtataatg attacgaaga attgaaacat 1500
tcaagtttgc ttaataggaa acaagaattg ttgaaagaat gtgaatcaaa taaaggacat 1560
atggatatgg agttgaggga gtttgaacaa tttttggtag atgctgtagg taactaa 1617

<210> 2
<211> 538
<212> PRT
<213> Candida albicans

<400> 2

```
Met Ser Asn Asn Ile Lys Val Ser Val Arg Val Arg Pro Leu Leu Pro
 1               5                  10                  15
Arg Glu Leu Lys Ala Lys Lys Ala Glu Gly Glu Ala Thr Pro Val Ser
            20                  25                  30
Leu Val Ser Met Pro Pro Ser Asp Pro His Lys Val Leu Leu Thr Asn
        35                  40                  45
Pro Lys Ala Gln Ser Lys Asn Asn Ile Ser Ile Asp Asn Lys Pro Lys
    50                  55                  60
Ser Phe Leu Phe Asp Glu Cys Ile Trp Ser Phe Asn Ser Lys Asp Ser
65                  70                  75                  80
Asn Tyr Thr Asp Asn His Lys Tyr Tyr Glu Lys Thr Gly Pro Gln Leu
                85                  90                  95
Leu Ser His Phe Phe Glu Gly Tyr Asn Val Cys Leu Leu Ala Tyr Gly
            100                 105                 110
Gln Thr Gly Ser Gly Lys Thr Tyr Thr Met Met Gly Asp Lys Asn Glu
        115                 120                 125
Ser Pro Gly Ile Ile Pro Leu Leu Ile Lys Asp Ile Leu Lys Gln Met
    130                 135                 140
Glu Ile Ser Val Asn Glu Lys Ile Asn Cys Glu Leu Lys Phe Ser Tyr
145                 150                 155                 160
Ile Glu Ile Tyr Asn Glu Gln Val Lys Asp Leu Leu Thr Gly Asp Thr
                165                 170                 175
Lys Asn Gly Asn Gly Asn Gly Asn Ser Thr His Ala Ser Gln Lys Cys
            180                 185                 190
Lys Val Arg Glu His Pro Val Thr Gly Pro Tyr Val Glu Asn Val Lys
        195                 200                 205
Glu Tyr Asn Ile Glu Asn Tyr Thr Gln Phe Leu Lys Leu Leu Ala Lys
    210                 215                 220
Gly Asn Ser Ser Arg Ser Thr Ala Ser Thr Ser Met Asn Asp Lys Ser
225                 230                 235                 240
Ser Arg Ser His Ala Ile Ile Thr Leu Thr Leu Lys Gln Thr Lys Phe
                245                 250                 255
Glu Lys Phe Ser Gly Phe Ser Ser Asp Ser Asp Glu Thr Asn Ser Ile
            260                 265                 270
Gly Asp Ala Ser Glu Glu Met Val Ser Asn Ile Lys Leu Val Asp Leu
        275                 280                 285
Ala Gly Ser Glu Arg Leu Gln Lys Thr Lys Leu Tyr Gly Gln Gln Glu
    290                 295                 300
Arg Ile Lys Glu Gly Thr Leu Ile Asn Lys Ser Leu Ala Val Leu Gly
305                 310                 315                 320
Arg Cys Ile Asn Leu Leu Ser Ser Asn Pro Ser Ser Leu Ile Pro Tyr
                325                 330                 335
Arg Glu Ser Ile Leu Thr Tyr Leu Leu Lys Glu Asn Leu Gly Gly Asn
            340                 345                 350
Ser Lys Thr Cys Met Ile Phe Cys Val Ser Pro Val Asp Tyr Glu Glu
        355                 360                 365
Thr His Gln Thr Leu Asn Tyr Ala Asn Glu Val Lys Lys Ile Lys Thr
    370                 375                 380
Leu Ala Lys Ala Asn Lys Thr Lys Leu Ser Asn Val Pro Ile Asp Trp
385                 390                 395                 400
Gln Gln Leu Gln Ser Thr Asp Lys Thr Val Ile Gln Ser Leu Lys Asn
                405                 410                 415
Glu Ile Asp Val Leu Thr Asn Lys Leu Asn Gln Met Gln Gln Leu Asn
            420                 425                 430
Glu Pro Arg Ser His Pro Pro Gln Ser Ile Asn Asn Ile Ile Glu Tyr
        435                 440                 445
Leu Asp Lys Glu Thr Asn Lys Val Lys Phe Glu Asn Lys Tyr Leu Lys
    450                 455                 460
Gln Met Met Asn Thr Lys Asn Asn His Ile Lys Gly Leu Asn Asn His
```

```
465                 470                 475                 480
Ile Asp Phe Ile Ser Asn Glu Tyr Gln Leu Leu Tyr Asn Asp Tyr Glu
                485                 490                 495
Glu Leu Lys His Ser Ser Leu Leu Asn Arg Lys Gln Glu Leu Leu Lys
            500                 505                 510
Glu Cys Glu Ser Asn Lys Gly His Met Asp Met Glu Leu Arg Glu Phe
        515                 520                 525
Glu Gln Phe Leu Val Asp Ala Val Gly Asn
    530                 535
```

<210> 3
<211> 1146
<212> DNA
<213> Candida Albicans

<400> 3

```
tcaaacaata tcaaagtatc agttcgagtt agacctttat tgccacggga gctaaaagct 60
aaaaaggccg aaggggaggc cacaccagta tcattagtgt caatgccccc gtcagatcct 120
cataaagttc tattgaccaa ccccaaggca caatcgaaga acaatatcag tattgacaac 180
aagcccaaga gttttttatt tgatgagtgt atttggtcat ttaatagtaa ggattcgaat 240
tacactgata atcacaaata ttatgaaaaa accggaccac aattgttgtc ccatttcttt 300
gaagggtata atgtttgcct tttggcgtat ggtcaaacag gatcagggaa aacatataca 360
atgatgggtg ataaaaatga atcgccaggg ataataccat tgttgataaa agatatctta 420
aagcaaatgg agatttctgt taatgagaaa atcaattgtg aattgaagtt ttcgtacatt 480
gaaatataca atgagcaagt caaagatttg cttacaggtg acaccaaaaa tgggaacggg 540
aatgggaaca gtacacatgc aagccaaaag tgcaaagtta gagaacatcc tgttactggt 600
ccatacgtcg agaatgtgaa agaatacaat atcgaaaatt atacccaatt tttgaaattg 660
ttggctaaag gtaattctag tcgttcaaca gcatctacat ctatgaacga taaaagttca 720
agaagtcacg caattatcac attgacatta aagcaaacca aatttgaaaa atttagcggt 780
tttagcagtg attctgatga aaccaacagt attggtgatg caagtgaaga gatggtatcg 840
aacatcaagt tggtcgatct tgctggatca gaaagactac agaaaacaaa attatatgga 900
caacaagaaa gaatcaaaga agggacattg attaacaaat cattggctgt tttgggaaga 960
tgtataaatt tgctttcaag caatccatcg tcattgatcc cctaccgtga atcaatttta 1020
acatatttat taaaagaaaa tcttggtggg aattctaaaa catgcatgat cttttgtgtt 1080
tctccagttg actatgaaga aacccatcaa acattgaact acgccaatga agtgaaaaaa 1140
atcaaa                                                          1146
```

<210> 4
<211> 382
<212> PRT
<213> Candida Albicans

<400> 4

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

```
Pro Gly Ile Ile Pro Leu Leu Ile Lys Asp Ile Leu Lys Gln Met Glu
    130                 135                 140
Ile Ser Val Asn Glu Lys Ile Asn Cys Glu Leu Lys Phe Ser Tyr Ile
145                 150                 155                 160
Glu Ile Tyr Asn Glu Gln Val Lys Asp Leu Leu Thr Gly Asp Thr Lys
                165                 170                 175
Asn Gly Asn Gly Asn Gly Asn Ser Thr His Ala Ser Gln Lys Cys Lys
            180                 185                 190
Val Arg Glu His Pro Val Thr Gly Pro Tyr Val Glu Asn Val Lys Glu
        195                 200                 205
Tyr Asn Ile Glu Asn Tyr Thr Gln Phe Leu Lys Leu Leu Ala Lys Gly
    210                 215                 220
Asn Ser Ser Arg Ser Thr Ala Ser Thr Ser Met Asn Asp Lys Ser Ser
225                 230                 235                 240
Arg Ser His Ala Ile Ile Thr Leu Thr Leu Lys Gln Thr Lys Phe Glu
                245                 250                 255
Lys Phe Ser Gly Phe Ser Ser Asp Ser Asp Glu Thr Asn Ser Ile Gly
            260                 265                 270
Asp Ala Ser Glu Glu Met Val Ser Asn Ile Lys Leu Val Asp Leu Ala
        275                 280                 285
Gly Ser Glu Arg Leu Gln Lys Thr Lys Leu Tyr Gly Gln Gln Glu Arg
    290                 295                 300
Ile Lys Glu Gly Thr Leu Ile Asn Lys Ser Leu Ala Val Leu Gly Arg
305                 310                 315                 320
Cys Ile Asn Leu Leu Ser Ser Asn Pro Ser Ser Leu Ile Pro Tyr Arg
                325                 330                 335
Glu Ser Ile Leu Thr Tyr Leu Leu Lys Glu Asn Leu Gly Gly Asn Ser
            340                 345                 350
Lys Thr Cys Met Ile Phe Cys Val Ser Pro Val Asp Tyr Glu Glu Thr
        355                 360                 365
His Gln Thr Leu Asn Tyr Ala Asn Glu Val Lys Lys Ile Lys
    370                 375                 380
```

<210> 5
<211> 1617
<212> DNA
<213> Candida Albicans

<400> 5

```
atgtcaaaca atatcaaagt atcagttcga gttagacctt tattgccacg ggagctaaaa 60
gctaaaaagg ccgaagggga ggccacacca gtatcattag tgtcaatgcc cccgtcagat 120
cctcataaag ttctattgac caaccccaag gcacaatcga agaacaatat cagtattgat 180
aacaagccca agagtttttt atttgatgag tgtatttggt catttaatag taaggattcg 240
aattacattg ataatcacaa atattatgaa aaaaccggac cacaattgtt gtcccatttc 300
tttgaagggt ataatgtttg ccttttggcg tatggtcaaa caggatcagg gaaaacatat 360
acaatgatgg gtgataaaaa tgaatcgcca gggataatac cattgttgat aaaagatatc 420
ttaaagcaaa tggagatttc tgttaatgag aaaatcaatt gtgaattgaa gttttcgtac 480
attgaaatat acaatgagca agtcaaagat ttgcttacag gtgacaccaa aaatgggaac 540
gggaatggga acagtacaca tgcaagccaa aagtgcaaag ttagagaaca tcctgttact 600
ggtccatacg tcgagaatgt gaaagaatac aatatcgaaa attataccca atttttgaaa 660
ttgttggcta aaggtaattc tagtcgttca acagcatcta catctatgaa cgataaaagt 720
tcaagaagtc acgcaattat cacattgaca ttaaagcaaa ccaaatttga aaaatttagc 780
ggttctagca gtgattctga tgaaaccaac agtattggtg atgcaagtga agagatggta 840
tcgaacatca agttggtcga tcttgctgga tcagaaagac tacagaaaac aaaattatat 900
ggacaacaag aaagaatcaa agaagggaca ttgattaaca aatcattggc tgttttggga 960
agatgtataa atttgctttc aagcaatcca tcgtcattga tcccctaccg tgaatcaatt 1020
ttaacatatt tattaaaaga aaatcttggt gggaattcta aaacatgcat gatcttttgt 1080
gtttctccag ttgactatga agaaacccat caaacattga actacgccaa tgaagtgaaa 1140
aaaatcaaaa cactagccaa ggcaaataaa acaaaacttt caaatgttcc tattgattgg 1200
```

caacagttac aatcgacaga taaaacagtt atccaatctt tgaagaatga aattgacgtt 1260
ttaacaaata aattaaatca aatgcaacaa ttaaatgaac ctagatctca tccaccacaa 1320
tcaataaata atataattga atatttggat aaagaaacta ataaagtgaa atttgaaaat 1380
aaatatttaa aacaaatgat gaataccaaa aacaatcaca taaaaggatt gaacaatcac 1440
attgatttta tttcgaacga atatcaattg ttgtataatg attacgaaga attgaaacat 1500
tcaagtttgc ttaataggaa acaagaattg ttgaaagaat gtgaatcaaa taaaggacat 1560
atggatatgg agttgaggga gtttgaacaa ttttttggtag atgctgtagg taactaa 1617

<210> 6
<211> 538
<212> PRT
<213> Candida albicans

<400> 6

Met	Ser	Asn	Asn	Ile	Lys	Val	Ser	Val	Arg	Val	Arg	Pro	Leu	Leu	Pro
1				5				10				15			
Arg	Glu	Leu	Lys	Ala	Lys	Lys	Ala	Glu	Gly	Glu	Ala	Thr	Pro	Val	Ser
		20					25					30			
Leu	Val	Ser	Met	Pro	Pro	Ser	Asp	Pro	His	Lys	Val	Leu	Leu	Thr	Asn
	35					40					45				
Pro	Lys	Ala	Gln	Ser	Lys	Asn	Asn	Ile	Ser	Ile	Asp	Asn	Lys	Pro	Lys
50					55					60					
Ser	Phe	Leu	Phe	Asp	Glu	Cys	Ile	Trp	Ser	Phe	Asn	Ser	Lys	Asp	Ser
65				70					75						80
Asn	Tyr	Ile	Asp	Asn	His	Lys	Tyr	Tyr	Glu	Lys	Thr	Gly	Pro	Gln	Leu
			85					90					95		
Leu	Ser	His	Phe	Phe	Glu	Gly	Tyr	Asn	Val	Cys	Leu	Leu	Ala	Tyr	Gly
		100					105					110			
Gln	Thr	Gly	Ser	Gly	Lys	Thr	Tyr	Thr	Met	Met	Gly	Asp	Lys	Asn	Glu
	115					120					125				
Ser	Pro	Gly	Ile	Ile	Pro	Leu	Leu	Ile	Lys	Asp	Ile	Leu	Lys	Gln	Met
130					135					140					
Glu	Ile	Ser	Val	Asn	Glu	Lys	Ile	Asn	Cys	Glu	Leu	Lys	Phe	Ser	Tyr
145				150					155						160
Ile	Glu	Ile	Tyr	Asn	Glu	Gln	Val	Lys	Asp	Leu	Leu	Thr	Gly	Asp	Thr
			165					170					175		
Lys	Asn	Gly	Asn	Gly	Asn	Gly	Asn	Ser	Thr	His	Ala	Ser	Gln	Lys	Cys
		180					185					190			
Lys	Val	Arg	Glu	His	Pro	Val	Thr	Gly	Pro	Tyr	Val	Glu	Asn	Val	Lys
	195					200					205				
Glu	Tyr	Asn	Ile	Glu	Asn	Tyr	Thr	Gln	Phe	Leu	Lys	Leu	Leu	Ala	Lys
210					215					220					
Gly	Asn	Ser	Ser	Arg	Ser	Thr	Ala	Ser	Thr	Ser	Met	Asn	Asp	Lys	Ser
225				230					235						240
Ser	Arg	Ser	His	Ala	Ile	Ile	Thr	Leu	Thr	Leu	Lys	Gln	Thr	Lys	Phe
			245					250					255		
Glu	Lys	Phe	Ser	Gly	Ser	Ser	Ser	Asp	Ser	Asp	Glu	Thr	Asn	Ser	Ile
		260					265					270			
Gly	Asp	Ala	Ser	Glu	Glu	Met	Val	Ser	Asn	Ile	Lys	Leu	Val	Asp	Leu
	275					280					285				
Ala	Gly	Ser	Glu	Arg	Leu	Gln	Lys	Thr	Lys	Leu	Tyr	Gly	Gln	Gln	Glu
290					295					300					
Arg	Ile	Lys	Glu	Gly	Thr	Leu	Ile	Asn	Lys	Ser	Leu	Ala	Val	Leu	Gly
305				310					315						320
Arg	Cys	Ile	Asn	Leu	Leu	Ser	Ser	Asn	Pro	Ser	Ser	Leu	Ile	Pro	Tyr
			325					330					335		
Arg	Glu	Ser	Ile	Leu	Thr	Tyr	Leu	Leu	Lys	Glu	Asn	Leu	Gly	Gly	Asn
		340					345					350			
Ser	Lys	Thr	Cys	Met	Ile	Phe	Cys	Val	Ser	Pro	Val	Asp	Tyr	Glu	Glu
	355					360					365				

```
Thr His Gln Thr Leu Asn Tyr Ala Asn Glu Val Lys Lys Ile Lys Thr
    370                 375                 380
Leu Ala Lys Ala Asn Lys Thr Lys Leu Ser Asn Val Pro Ile Asp Trp
385                 390                 395                 400
Gln Gln Leu Gln Ser Thr Asp Lys Thr Val Ile Gln Ser Leu Lys Asn
                405                 410                 415
Glu Ile Asp Val Leu Thr Asn Lys Leu Asn Gln Met Gln Gln Leu Asn
            420                 425                 430
Glu Pro Arg Ser His Pro Pro Gln Ser Ile Asn Asn Ile Ile Glu Tyr
        435                 440                 445
Leu Asp Lys Glu Thr Asn Lys Val Lys Phe Glu Asn Lys Tyr Leu Lys
    450                 455                 460
Gln Met Met Asn Thr Lys Asn Asn His Ile Lys Gly Leu Asn Asn His
465                 470                 475                 480
Ile Asp Phe Ile Ser Asn Glu Tyr Gln Leu Leu Tyr Asn Asp Tyr Glu
                485                 490                 495
Glu Leu Lys His Ser Ser Leu Leu Asn Arg Lys Gln Glu Leu Leu Lys
            500                 505                 510
Glu Cys Glu Ser Asn Lys Gly His Met Asp Met Glu Leu Arg Glu Phe
        515                 520                 525
Glu Gln Phe Leu Val Asp Ala Val Gly Asn
    530                 535
```

<210> 7
<211> 1146
<212> DNA
<213> Candida albicans

<400> 7

```
tcaaacaata tcaaagtatc agttcgagtt agacctttat tgccacggga gctaaaagct 60
aaaaaggccg aaggggaggc cacaccagta tcattagtgt caatgccccc gtcagatcct 120
cataaagttc tattgaccaa ccccaaggca caatcgaaga acaatatcag tattgataac 180
aagcccaaga gtttttttatt tgatgagtgt atttggtcat ttaatagtaa ggattcgaat 240
tacattgata atcacaaata ttatgaaaaa accggaccac aattgttgtc ccatttcttt 300
gaagggtata atgtttgcct tttggcgtat ggtcaaacag gatcagggaa aacatataca 360
atgatgggtg ataaaaatga atcgccaggg ataataccat tgttgataaa agatatctta 420
aagcaaatgg agatttctgt taatgagaaa atcaattgtg aattgaagtt ttcgtacatt 480
gaaatataca atgagcaagt caaagatttg cttacaggtg acaccaaaaa tgggaacggg 540
aatgggaaca gtacacatgc aagccaaaag tgcaaagtta gagaacatcc tgttactggt 600
ccatacgtcg agaatgtgaa agaatacaat atcgaaaatt atacccaatt tttgaaattg 660
ttggctaaag gtaattctag tcgttcaaca gcatctacat ctatgaacga taaaagttca 720
agaagtcacg caattatcac attgacatta aagcaaacca aatttgaaaa atttagcggt 780
tctagcagtg attctgatga aaccaacagt attggtgatg caagtgaaga gatggtatcg 840
aacatcaagt tggtcgatct tgctggatca gaaagactac agaaaacaaa attatatgga 900
caacaagaaa gaatcaaaga agggacattg attaacaaat cattggctgt tttgggaaga 960
tgtataaatt tgctttcaag caatccatcg tcattgatcc cctaccgtga atcaatttta 1020
acatatttat taaaagaaaa tcttggtggg aattctaaaa catgcatgat cttttgtgtt 1080
tctccagttg actatgaaga aacccatcaa acattgaact acgccaatga agtgaaaaaa 1140
atcaaa                                                          1146
```

<210> 8
<211> 382
<212> PRT
<213> Candida albicans

<400> 8

```
Ser Asn Asn Ile Lys Val Ser Val Arg Val Arg Pro Leu Leu Pro Arg
 1               5                  10                  15
Glu Leu Lys Ala Lys Lys Ala Glu Gly Glu Ala Thr Pro Val Ser Leu
```

```
            20                  25                  30
Val Ser Met Pro Pro Ser Asp Pro His Lys Val Leu Leu Thr Asn Pro
        35                  40                  45
Lys Ala Gln Ser Lys Asn Asn Ile Ser Ile Asp Asn Lys Pro Lys Ser
    50                  55                  60
Phe Leu Phe Asp Glu Cys Ile Trp Ser Phe Asn Ser Lys Asp Ser Asn
65                  70                  75                  80
Tyr Ile Asp Asn His Lys Tyr Tyr Glu Lys Thr Gly Pro Gln Leu Leu
                85                  90                  95
Ser His Phe Phe Glu Gly Tyr Asn Val Cys Leu Leu Ala Tyr Gly Gln
            100                 105                 110
Thr Gly Ser Gly Lys Thr Tyr Thr Met Met Gly Asp Lys Asn Glu Ser
        115                 120                 125
Pro Gly Ile Ile Pro Leu Leu Ile Lys Asp Ile Leu Lys Gln Met Glu
    130                 135                 140
Ile Ser Val Asn Glu Lys Ile Asn Cys Glu Leu Lys Phe Ser Tyr Ile
145                 150                 155                 160
Glu Ile Tyr Asn Glu Gln Val Lys Asp Leu Leu Thr Gly Asp Thr Lys
                165                 170                 175
Asn Gly Asn Gly Asn Gly Asn Ser Thr His Ala Ser Gln Lys Cys Lys
            180                 185                 190
Val Arg Glu His Pro Val Thr Gly Pro Tyr Val Glu Asn Val Lys Glu
        195                 200                 205
Tyr Asn Ile Glu Asn Tyr Thr Gln Phe Leu Lys Leu Leu Ala Lys Gly
    210                 215                 220
Asn Ser Ser Arg Ser Thr Ala Ser Thr Ser Met Asn Asp Lys Ser Ser
225                 230                 235                 240
Arg Ser His Ala Ile Ile Thr Leu Thr Leu Lys Gln Thr Lys Phe Glu
                245                 250                 255
Lys Phe Ser Gly Ser Ser Ser Asp Ser Asp Glu Thr Asn Ser Ile Gly
            260                 265                 270
Asp Ala Ser Glu Glu Met Val Ser Asn Ile Lys Leu Val Asp Leu Ala
        275                 280                 285
Gly Ser Glu Arg Leu Gln Lys Thr Lys Leu Tyr Gly Gln Gln Glu Arg
    290                 295                 300
Ile Lys Glu Gly Thr Leu Ile Asn Lys Ser Leu Ala Val Leu Gly Arg
305                 310                 315                 320
Cys Ile Asn Leu Leu Ser Ser Asn Pro Ser Ser Leu Ile Pro Tyr Arg
                325                 330                 335
Glu Ser Ile Leu Thr Tyr Leu Leu Lys Glu Asn Leu Gly Gly Asn Ser
            340                 345                 350
Lys Thr Cys Met Ile Phe Cys Val Ser Pro Val Asp Tyr Glu Glu Thr
        355                 360                 365
His Gln Thr Leu Asn Tyr Ala Asn Glu Val Lys Lys Ile Lys
    370                 375                 380
```