METHODS FOR ANALYZING LARIAT RNA

The present invention relates to compositions and methods useful for analyzing lariat RNA, which plays a role in the regulation of gene expression. A sample of RNA is specifically treated to remove linear miRNA and enrich for lariat RNA. The enriched lariat RNA sample may be analyzed further to identify introns, branch point sequences, alternative splicing patterns, and gene transcription levels. The enriched lariat RNA sample may also be exploited as a detection or compound screening tool, as well as other uses.
FIG. 1A

ACT1 gene

pre-mRNA

mRNA

intron

2'→5' bond

FIG. 1B

15 cycles

11 cycles

wt  dbr1  wt  dbr1

mRNA  intron  mRNA  intron  mRNA  intron  mRNA  intron

-215/216

-146/363
FIG. 3A

FLO8 mRNA (2400 nt)

FIG. 3B

ACT1 1/2 3/4 5/6 7/8 9/10 11/12

M - + - + - + - + - + - + - +

FIG. 3C

ACT1 1/2 3/4 5/6 7/8 9/10 11/12

M - + - + - + - + - + - + - +
FIG. 5

Dbr1p: - - + +

-mRNA
- lariat

M 1 2 3 4
FIG. 6A

<table>
<thead>
<tr>
<th>PNPase</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dbr1p</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

M 1 2 3 4 5 6 7 8

FIG. 6B

<table>
<thead>
<tr>
<th>Dbr1p</th>
<th>-</th>
<th>-</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNPase</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

M 1 2 3 4

m 5 6 7 8
FIG. 7C

Relative *RPP1B* intron lariat levels

FIG. 7D

Relative *YRA1* intron lariat levels
FIG. 11

<table>
<thead>
<tr>
<th>lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>mRNA</td>
<td>intron</td>
<td>mRNA</td>
<td>intron</td>
<td>mRNA</td>
<td>intron</td>
</tr>
</tbody>
</table>

Step 1 ±Dbr1p: - - + + - -
Step 2 ±PNPase: + + + + - -

FIG. 12

<table>
<thead>
<tr>
<th>lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>mRNA</td>
<td>intron</td>
<td>mRNA</td>
<td>intron</td>
</tr>
</tbody>
</table>

Dbr1p: - - + +
FIG. 13B

ACT1 intron 57 seq reads

Number of sequence reads

Chromosome position (kbp)
METHODS FOR ANALYZING LARIAT RNA

FIELD OF THE INVENTION

[0001] The present invention relates to compositions, methods, and kits for analyzing lariat RNA. In particular, the invention relates to enriching an RNA population for lariat RNA and then analyzing the lariat RNA population.

BACKGROUND OF THE INVENTION

[0002] Pre-mRNA introns play an important role in the regulation of gene expression for many eukaryotes because their presence allows for the occurrence of alternative splicing. Such alternative splicing results in the creation of multiple proteins from a single gene, many of which are expressed in cell- or tissue-specific patterns. The pre-mRNA introns are excised in a lariat conformation to produce mRNA. Following excision, the 3' tails of the lariats are subject to exonucleolytic degradation up to the lariat branch point. The predominant pathway for further exonucleolytic degradation requires cleavage of the 2'-5' bond located at the branch point. This cleavage event occurs via a RNA debranching enzyme, a 2'-5' phosphodiesterase.

[0003] Although intron RNA sequences contain information necessary for their removal from pre-mRNAs, some introns contain additional information. In most eukaryotes microRNAs (miRNAs) and small nucleolar RNAs (snRNAs) are encoded within introns. In studies with human cells it has been found that the vast majority of intronic miRNAs are excised from pre-mRNAs. Intron snRNAs, on the other hand, are processed from excised introns, as determined in baker's yeast, humans, and other eukaryotes.

[0004] Debranching and subsequent degradation of most intron RNAs are rapid, resulting in low steady state levels of intron RNAs relative to the corresponding mature RNAs. The exceptions are intron sequences corresponding to RNAs with additional functions (e.g. snRNAs). Studies in many different organisms have determined that cleavage of the 2'-5' bond by an RNA debranching enzyme is important for the maturation of intron-encoded snRNAs and mirtrons, which is another class of miRNAs that are processed from excised introns.

[0005] Genome-wide studies analyzing excised intron RNAs in fruit flies and yeast have identified new introns and alternative splicing patterns. These analyses relied on creating cell populations that accumulate excised intron RNAs at elevated levels due to either mutation of the gene encoding debranching enzyme or knock down of debranching enzyme expression with siRNA. Analysis of RNA samples with elevated levels of RNA lariats increases the detectability of rare splicing variants. Cells defective for RNA debranching activity accumulate excised introns in their lariat forms with shorted 3'tails. Without the full length 3'tail, information for the 3' exon-junction is not obtainable from the intron lariat RNA sequences. However, studies have shown that the positions of RNA branch points may be deduced from analyzing intron RNA lariats. Direct information on branch points is only obtainable from analysis of RNA lariats. Therefore, there is a need to provide new compositions and methods for the analysis of RNA lariats that allow analysis of rare splicing variants and branch point sequences.

REFERENCE TO COLOR FIGURES

[0006] The application file contains at least one photograph executed in color. Copies of this patent application publication with color photographs will be provided by the Office upon request and payment of the necessary fee. The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0008] FIG. 1 illustrates reverse transcriptase polymerase chain reaction (RT-PCR) detection of lariat RNA. FIG. 1A shows the annealing positions of primers for RT-PCR detection of ACT1 gene intron lariat RNA and mRNA. The intron lariat RNA is detected using primers oligo 146 and oligo 365 (depicted by arrows within the intron lariat RNA loop). The linear mRNA is detected using primers oligo 215 and oligo 216 (depicted by small arrows below the mRNA arrow). FIG. 1B shows an agarose gel analysis of RT-PCRs for ACT1 RNA detected using the primers illustrated in FIG. 1A (mRNA=oligos 215/216; intron=oligos 146/365). Lanes 1-4 contain reactions run 15 cycles after the touchdown phase of polymerase chain reaction (PCR); lanes 5-8 contain reactions run 11 cycles after the touchdown phase of PCR; lanes 1, 2, 5, and 6 contain reactions using wild type (DBR1) RNA samples; and, lanes 3, 4, 7, and 8 contain reactions using dB1 mutant RNA samples. The different numbers of cycles were run to show the linearity of the PCR.

[0009] FIG. 2 illustrates selective degradation of linear RNAs and not lariat RNAs. FIG. 2A shows an agarose gel analysis of RT-PCRs for ACT1 intron lariat RNA following a series of enzyme treatments (PNPase) at decreasing amounts (indicated by the wedge at the top of the gel image, highest amount of enzyme used in lane 1 and lowest in lane 7). RT-PCRs for ACT1 intron lariat RNA were performed with primers 146 and 365 (see FIG. 1) and run for 15 cycles after the touchdown phase of the reaction.

[0010] FIG. 2B shows an agarose gel analysis of RT-PCRs for ACT1 linear mRNA from the same series of enzyme treatments of FIG. 2A. Lanes 15 and 16 contain ACT1 intron lariat and ACT1 linear mRNA RT-PCRs, respectively, of mRNA samples that did not undergo PNPase treatment. The products in these lanes serve as size markers for the intron lariat and mRNA products in lanes 1-14. RT-PCRs for ACT1 mRNA were performed with primers 215 and 216 (see FIG. 1) and run for 24 cycles after the touchdown phase of the reaction.

[0011] FIG. 3 illustrates processivity of PNPase on FLO8 mRNA. FIG. 3A shows primer pairs for amplifying different segments along the length of FLO8 mRNA: 1/2 primers 372 and 373; 3/4 primers 374 and 375; 5/6 primers 376 and 377; 7/8 primers 378 and 379; 9/10 primers 380 and 381; 11/12 primers 382 and 383. FIG. 3B shows a PAGE analysis of RT-PCRs for FLO8 mRNA segments following enzyme treatment (PNPase, +lanes) and mock treatment (–lanes) of a total cellular RNA sample that had been pretreated with DNase I. Lanes containing the various FLO8 RT-PCRs are indicated below the gel image; the FLO8 primer pairs are
indicated above the gel image. RT-PCRs for ACT1 RNAs are in the four lanes under the ACT1 title and serve as controls that indicate the PNPass reactions proceeded as expected. The RT-PCRs for ACT1 mRNA and intron RNA are indicated below the corresponding lanes. These reactions used primer pairs 215/216 and 146/363, respectively. The lane marked “M” contains a DNA molecular weight standard (50 base pair (bp) ladder). FIG. 3C shows a PAGE analysis of RT-PCRs as described for FIG. 3B except that the total cellular nucleic acid samples were not treated with DNase I prior to PNPass enzyme treatment and RT-PCRs. For all RT-PCRs in FIGS. 3B and 3C, reactions were performed with 24 cycles after the touchdown phase.

[0012] FIG. 4 illustrates the purification of Dbr1p. FIG. 4A shows the elution profile of histidine-tagged yeast Dbr1p, purified from E. coli, collected for 100 mM and 200 mM concentrations of imidazole. Dbr1p bound to a nickel-nitriotriacetate acid (nickel-NTA) column was eluted with increasing concentrations of imidazole. Six, ~1.5 mL fractions were collected for each imidazole concentration. FIG. 4B shows the elution profile of histidine-tagged yeast Dbr1p collected for 300 mM and 500 mM concentrations of imidazole. Key: “M” is the protein molecular weight standard; and, 1-6 are the fractions collected from the nickel-nitriotriacetate column. FIG. 4C shows the precipitation of histidine-tagged yeast Dbr1p collected for 300 mM and 500 mM concentrations of imidazole. Key: “M” is the protein molecular weight standard; and, 1-6 are the fractions collected from the nickel-nitriotriacetate column. FIG. 4D shows the precipitation of histidine-tagged yeast Dbr1p collected for 300 mM and 500 mM concentrations of imidazole. Key: “M” is the protein molecular weight standard; and, 1-6 are the fractions collected from the nickel-nitriotriacetate column.

[0013] FIG. 5 illustrates an in vitro debranching reaction. Specifically, FIG. 5 shows an agarose gel analysis of RT-PCRs for ACT1 RNAs following treatment with Dbr1p (+lanes) and mock treatment (-lanes) of total cellular RNA sample. Key: lanes 2 and 4 contain RT-PCRs for ACT1 intron lariat RNA; lanes 1 and 3 are RT-PCRs for ACT1 mRNA; lane M contains a DNA molecular weight standard (50 bp ladder). RT-PCRs for ACT1 intron lariat RNA were run for 14 cycles after the touchdown phase of the reaction; RT-PCRs for ACT1 mRNA were run for 24 cycles after the touchdown phase of the reaction.

[0014] FIG. 6 shows combinations of PNPass and Dbr1p enzyme treatments. FIG. 6A shows an agarose gel analysis of RT-PCRs for ACT1 RNAs following treatment of a total cellular RNA sample from a dbr1 strain with Dbr1p (+Dbr1p) and PNPass (+PNPass) as well as mock treatment (-treatment). In this experiment, PNPass treatment preceded Dbr1p treatment for samples that were treated with both enzymes. Lanes 1, 3, 5, and 7 contain RT-PCRs for ACT1 mRNA of a total cellular RNA sample. Lanes 2, 4, 6, and 8 contain parallel RT-PCRs for ACT1 intron lariat RNA. FIG. 6B shows an agarose gel analysis of RT-PCRs for ACT1 RNAs following treatment of a total cellular RNA sample from a dbr1 strain with Dbr1p and PNPass as well as mock treatment. In this experiment, Dbr1p treatment preceded PNPass treatment for samples that were treated with both enzymes. Lanes 1-4 contain RT-PCRs for ACT1 mRNA of a total cellular RNA sample. Lanes 5-8 contain parallel RT-PCRs for ACT1 intron lariat RNA. For both FIGS. 6A and 6B, RT-PCRs for ACT1 intron lariat RNA were run for 19 cycles after the touchdown phase of the reaction and RT-PCRs for ACT1 mRNA were run for 24 cycles after the touchdown phase of the reaction. The lanes marked “M” and “m” contain DNA molecular weight standards (“M”–A phage DNA cut with HindIII+EcoRI; “m”–50 bp ladder).

[0015] FIG. 7 shows real-time quantitative RT-PCR (qRT-PCR) measurement of lariat RNA levels. FIG. 7A shows the annealing positions of primers for RT-PCR detection of mRNA (FWDm primer and REVm primer) and intron lariat RNA species (FWDi primer and REVi primer). A TaqMan probe is designed to span the same exon-exon junction. The star and the triangle at opposite ends of the TaqMan probes represent the fluorescent reporter molecule and the quencher that are bound to the 5’ and 3’ ends, respectively. The TaqMan probes that anneal to a particular mRNA and lariat RNA pair contain different fluorescent reporter molecules, indicated by solid and stippled stars. Note that lariat RNA detection does not involve annealing of PCR primers or TaqMan probes across lariat branch points. FIG. 7B graphically illustrates the relative quantification of ACT1 intron lariat RNA in total RNA samples from different yeast strains. RQ, the relative quantification, is the ratio of intron RNA to mRNA for a particular sample relative to the ratio of intron RNA to mRNA for the DBR1 (wild-type) sample at the left end of the bar graph (which sets the RQ for DBR1 telIto1). Quantification experiments were repeated three times and the qPCRs were performed in triplicate each time. The standard error bars display the calculated maximum (RQmax) and minimum (RQmin) expression levels that represent standard error of the mean expression level (RQ value). FIG. 7C graphically illustrates the relative quantification of RPP1B intron lariat RNA for the same RNA samples presented in FIG. 7B. FIG. 7D graphically illustrates the relative quantification of YRA1 intron lariat RNA for the same RNA samples presented in FIG. 7B.

[0016] FIG. 8 graphically illustrates a time course of an in vitro debranching reaction.

[0017] FIG. 9 illustrates the RNA lariat enrichment following treatment of an RNA sample with a 3’ exonuclease. The parentheses at the left end of the linear RNA mean that these RNAs include both 5’ capped and 5’ uncapped species. The circular dot within the parentheses represents the cap. The arrow on the right side of the linear RNA represents the 3’ end. Dashed lines represent degradation.

[0018] FIG. 10 illustrates the RNA lariat enrichment following treatment of a decapped RNA sample with a 3’ exonuclease. Linear RNAs at the top, below the lariat RNA, are a mixture of 5’ capped and 5’ uncapped species. The circular dot at the left of the 5’ capped RNA represents the cap. The arrows on the right side of the linear RNAs represent the 3’ ends. Dashed lines represent degradation.

[0019] FIG. 11 illustrates RT-PCR detection of ACT1 mRNA (linear RNA) and intron (lariat RNA) in a total RNA sample from Saccharomyces cerevisiae cells following treatment with the 3’ exonuclease polynucleotide phosphorylase (PNPass) (lanes 1 and 2), debranching enzyme (Dbr1p) followed by PNPass (lanes 3 and 4), and no treatment (lanes 5 and 6).

[0020] FIG. 12 illustrates the RT-PCR detection of ACT1 mRNA (linear RNA) and intron (lariat RNA) in total RNA samples from dbr1 mutant yeast cells following Dbr1p treatment (lanes 3 and 4) or no treatment (lanes 1 and 2).

[0021] FIG. 13 illustrates the high-throughput sequencing of cDNAs representing PNPass-treated S. cerevisiae RNA. FIG. 13A shows chromosome 6 is depicted at the top, below which a 20 kilo-base pair (kbp) segment is highlighted (black bar), along with a detailed map of the genes lie within this segment. Gene open reading frames (ORFs) are indicated by red or blue bars, depending on which DNA strand of the
chromosome encodes the sense strand for each ORF (red for the upper strand, blue for the lower strand). FIG. 13B graphically illustrates the number of sequence reads that map within the 20 kbp segment. The ACT1 gene is the only gene in this 20 kbp segment that contains an intron, which is depicted as a white box within the blue ACT1 ORF.

The present invention is directed to compositions and methods for analyzing lariat RNA. The compositions of the invention include isolated enzymes and supportive buffers for efficient use of the isolated enzymes. The methods of the invention include methods of enriching an RNA population for lariat RNA and analyzing lariat RNA. The compositions and methods of the invention may be provided in a kit.

The enzymes of the invention include linear RNA degrading enzymes, 5' cap removing enzymes and debranching enzymes. Suitable linear RNA degrading enzymes include those capable of degrading linear RNA or mRNA. Such linear RNA degrading enzymes include, without limitation, exonucleases, 5' exonucleases, those with both 5' and 3' exonuclease activity, those known in the art or yet to be discovered, and combinations thereof.

Suitable 5' cap removing enzymes include those capable of degrading or excising the 5' cap of linear RNA or mRNA. Such enzymes include those commonly known in the art, such as Dcp1 or Dcp2, as well as those yet to be discovered, and combinations thereof.

Suitable debranching enzymes include those capable of degrading, excising, or cleaving the 2'-5' bond at the branch point of lariat RNA. Such enzymes include 2'-5' phosphodiesterases, such as DbR1, all those known in the art or yet to be discovered, and combinations thereof. Also, such enzymes include those encoding an amino acid sequence having at least 35% sequence identity to at least one of SEQ ID NOs: 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66. The sequence identity may be about 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more. In another embodiment, the nucleic acid sequence may have at least 35% sequence identity to the metallophosphatase domain of at least one of SEQ ID NO: 46-66. The sequence identity may be about 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more. In S. cerevisiae Dbr1 (SEQ ID NO: 47), the metallophosphatase domain is located at amino acid residues 6 to 238 (FIG. 14).

The invention also includes methods of enriching an RNA population for lariat RNA. Such methods include providing an RNA population and contacting the RNA population with a linear RNA degrading enzyme to form a lariat RNA enriched population. Suitable methods may further include contacting the RNA population with a debranching enzyme.

The invention also includes methods of analyzing the lariat RNA in an RNA sample or population. Such methods include providing an RNA population and contacting the RNA population with a linear RNA degrading enzyme to form a lariat RNA enriched population. The lariat RNA enriched population may be used to create a cDNA library. In one embodiment the cDNA library is created by reverse transcribing the lariat RNA enriched population. Methods known in the art for creating a cDNA library may be used. Suitable methods may also further include sequencing the cDNA library created using the lariat RNA enriched population.

The invention includes kits for practicing the methods of the invention. Suitable kits contain at least one linear RNA degrading enzyme and instructions. Kits may also include a linear RNA degrading enzyme buffer, debranching enzyme, debranching enzyme buffer, 5' decapping enzyme, 5' decapping enzyme buffer, and combinations thereof.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In accordance with the present invention, processes of comprehensively analyzing lariat RNA have been discovered. In particular, the present invention provides compositions, methods, and kits useful for analyzing lariat RNA. The compositions and methods are directed to enriching an RNA population for lariat RNA and analyzing the lariat RNA.

Various aspects of the invention are described in further detail in the following subsections.

I. Compositions

A. Enzymes

One aspect of the invention pertains to isolated enzymes that are used in the methods described herein. Suitable enzymes include those capable of degrading linear RNA, linearizing lariat RNA, removing the 5' cap from linear RNA (mRNA), or combinations thereof.

Enzymes capable of degrading linear RNA are used to remove the linear RNA from the RNA population, enriching the population for lariat RNA. Suitable linear RNA degrading enzymes include, without limitation, 5' exonucleases, 5' exonuclease, and combinations thereof. Any enzyme capable of degrading linear RNA is contemplated herein, as well as those not yet discovered. For example, the polynucleotide phosphorylases of Bacillus stearothermophilus (BspNPase) and Thermus thermophilus (TspNPase), as well as the RNase of E. coli (RNase R) are suitable linear RNA degrading enzymes.

Enzymes capable of removing the 5' cap from linear RNA or mRNA are used to allow linear RNA degrading enzymes to work, where the 5' cap may inhibit degradation. Suitable 5' cap removing enzymes include those capable of cleaving or degrading the 5' cap from linear RNA or mRNA. Any enzyme capable of 5' cap removal is contemplated herein, as well as those not yet discovered. For example, the 5' cap removing enzymes Dcp1 and Dcp2 are suitable for the invention. The invention also includes 5' cap removal treatments known in the art or yet to be discovered.

Enzymes capable of linearizing lariat RNA are debranching enzymes, which are used to unfold the lariat structure of the RNA to allow further analysis. Suitable debranching enzymes are those capable of cleaving the 2'-5' bond at the branch point of lariat RNA. Such debranching
enzymes include, without limitation, debranching enzymes having sequence homology to SEQ ID NO: 46-66.

[0037] Preferably, the nucleic acid sequence of debranching enzymes have at least 35% sequence identity to the nucleic acid sequence that encodes the amino acid sequence of at least one of SEQ ID NO: 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66. The sequence identity may be about 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more. In another embodiment, the nucleic acid sequence may have at least 35% sequence identity to the metalophosphatase domain of the nucleic acid sequence that encodes at least one of SEQ ID NO: 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, or 66. The sequence identity may be about 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more. In S. cerevisiae Dbr1 (SEQ ID NO: 47), the metalophosphatase domain is located at amino acid residues 6 to 238.

[0038] A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 46-66, or a complement of any of these nucleotide sequences, may be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequences of SEQ ID NO:46-66, debranching enzyme nucleic acid molecules may be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

[0039] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding debranching enzymes that contain changes in amino acid residues that may or may not be essential for activity. Such debranching enzymes proteins differ in amino acid sequence from SEQ ID NO: 46-66. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to the amino acid sequence of SEQ ID NO: 46-66. An isolated nucleic acid molecule encoding a debranching enzymes having a sequence which differs from that of SEQ ID NO: 46-66, may be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of debranching enzymes (SEQ ID NO: 46-66) such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations may be introduced by standard techniques known in the art, such as site-directed mutagenesis and PCR-mediated mutagenesis.

[0040] The present invention encompasses antisense nucleic acid molecules. Antisense molecules are complementary to a sense nucleic acid encoding a protein, complementary to the coding strand of a double-stranded cDNA molecule, or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid hydrogen bonds to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire debranching enzyme coding strand, or to only a portion thereof, such as all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to a non-coding region of the coding strand of a nucleotide sequence encoding a debranching enzyme. The non-coding regions (‘5’ and ‘3’ untranslated regions”) are the 5’ and ‘3’ sequences that flank the coding region and are not translated into amino acids. Given the coding strand sequences encoding debranching enzymes disclosed herein, antisense nucleic acids of the invention may be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule may be complementary to the entire coding region of debranching enzyme mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or non-coding region of a debranching enzyme mRNA. For example, the antisense oligonucleotide may be complementary to the region surrounding the translation start site of a debranching enzyme mRNA. An antisense oligonucleotide may be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention may be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) may be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which may be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acyctelytosine, 5-carboxyhydroxymethyl uracil, 5-carboxyhexamethylaminoethyl 2-thiouridine, 5-carboxyaminomethyluracil, dihydrouracil, beta-D-galac tosyllactosone, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylthioadenine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylpyosine, 5’-methoxy thymoxymethyluracil, 5-methoxyuracil, 2-methylthio-N-6-isopentenyladenine, uracil-5-oxycetid acid (v), wybutosine, pseudouracil, queosine, 2-thiouracil, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methyl ester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid may be produced biologically using an expression vector into which a nucleic acid has been subcloned in in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest).

[0041] The antisense nucleic acid molecules of the invention are generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a debranching enzyme to thereby inhibit expression of the enzyme, e.g., by inhibiting transcription and/or translation. The hybridization may be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix.

[0042] The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988)
may be used to catalytically cleave debranching enzyme mRNA transcripts to thereby inhibit translation of debranching enzyme mRNA. A ribosome having specificity for a debranching enzyme-encoding nucleic acid may be designed based upon the nucleotide sequence of the debranching enzyme cDNA. For example, debranching enzyme mRNA may be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak (1993) Science 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, debranching enzyme gene expression may be inhibited by targeting nucleotide sequences complementary to the regulatory region of the debranching enzyme gene (e.g., promoter and/or enhancers) to form triple helical structures that prevent transcription of the debranching enzyme gene in target cells. See generally, Helene (1991) Anticancer Drug Des. 6(6):569-84; Helene (1992) Ann. N.Y. Acad. Sci. 660:27-36; and Maher (1992) Bioreas 14(12):807-15.

A useful debranching enzyme protein is a protein which includes an amino acid sequence at least about 45%, preferably 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or more identical to the amino acid sequence of SEQ ID NO: 46-66, and retains the functional activity of a debranching protein of SEQ ID NO: 46-66.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino acid or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity= # of identical positions/total # of positions x 100).

The determination of percent homology between two sequences may be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) Proc. Nat’l Acad. Sci. USA 87:2264-2268, modified as in Karlin and Altschul (1993) Proc. Nat’l Acad. Sci. USA 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) J. Mol. Biol. 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences similar or homologous to nucleic acid sequences of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

One useful fusion protein is a fusion protein in which the debranching enzyme sequences are fused to tag or marker sequences. Such fusion proteins can facilitate the purification of recombinant debranching enzymes. Suitable tag and marker sequences are well known in the prior art and include all those known in the art or yet to be discovered. Exemplary tags or markers include, without limitation, HIS tag, GST, MYC tag, fluorescent proteins, fluorophores, and others that are too numerous to include herein.

One skilled in the art will recognize that activity of enzymes depends upon conditions that are specific to each enzyme. Some enzymes are active at higher temperatures, such as 65°C, while others are active at lower temperatures, such as 37°C. Other conditions include pH and salt content. As such conditions depend upon the enzyme, the invention includes all conditions for which the enzymes useful for the invention are active.

II. Methods

The present invention includes methods of preparing and analyzing lariat RNA populations. Methods of the invention also include using the compositions described herein to modulate the proportion of lariat RNA in an RNA population.

Methods of preparing lariat RNA populations or enriched lariat RNA populations include providing an RNA population and contacting it with a linear RNA degrading enzyme to form a lariat RNA enriched population. In some embodiments, methods may further include contacting the RNA population with a debranching enzyme. The order with which the RNA population is contacted with the linear RNA degrading enzyme and debranching enzyme determines the composition of the resulting enriched RNA population. If the RNA population is contacted with the linear RNA degrading enzyme before the debranching enzyme, then the resulting enriched RNA population will be enriched for lariat RNA. If the RNA population is contacted with the debranching enzyme before the linear RNA degrading enzyme, then the resulting enriched RNA population will not be enriched for lariat RNA or linear RNA.

In some embodiments, methods may further include contacting the RNA population with a 5′ cap removing enzyme or be subjected to a 5′ cap removal treatment. Preferably, the 5′ cap removing enzyme or treatment is contacted or used on the RNA population before the linear RNA degrading enzyme.

In some embodiments, methods may include inhibiting the RNA debranching enzyme in a population of cells prior to the methods of enriching for lariat RNA. Inhibiting the RNA debranching enzyme in a population of cells would allow the proportion of lariat RNA in a population of cells to increase, thereby allowing the enriched lariat RNA population to increase. The RNA debranching enzyme may be inhibited using methods known in the art. Such methods may include, without limitation, siRNA technology, ribozymes, knockout cell lines, knock down cell lines, and other methods known in the art.

The invention also includes methods of analyzing the lariat RNA in an RNA sample or population. In some embodiments, methods include providing an RNA population and contacting the RNA population with a linear RNA degrading enzyme to form a lariat RNA enriched population. The lariat RNA enriched population is contacted with a debranching enzyme and then subsequently with a linear RNA degrading enzyme to confirm true lariat RNAs are present.

In other embodiments, methods include providing an RNA population and contacting the RNA population with a linear RNA degrading enzyme to form a lariat RNA
enriched population. The lariat RNA enriched population is then used to create a cDNA library. In one embodiment, the cDNA library is created by reverse transcribing the lariat RNA enriched population. Methods known in the art for creating a cDNA library may be used. Suitable methods may also further include sequencing the cDNA library created using the lariat RNA enriched population. Methods known in the art for sequencing may be used.

III. Kits

[0055] The present invention includes articles of manufacture and kits containing materials useful for preparing enriched lariat RNA populations as described herein. The article of manufacture may include a container of a composition as described herein with a label. Suitable containers include, for example, bottles, vials, and test tubes. The containers may be formed from a variety of materials such as glass or plastic.

[0056] In one embodiment, containers hold a composition having an active agent which is effective for degrading linear RNA or linearizing lariat RNA. The active agent may be an enzyme. Suitable enzymes include 3’ exonucleases, 5’ exonucleases, 5’/3’ exonucleases, debranching enzymes, decapping enzymes, or combinations thereof. Active agents may be combined into a single container or provided in separate containers. Preferably, the active agents are provided in separate containers.

[0057] In another embodiment, containers may hold a composition having a supportive agent, which is supportive of the active agent. Such supportive agents may be buffers. The supportive agent will depend upon the active agent. Exemplary supportive agents include, without limitation, exomuclease reaction buffer, debranching enzyme reaction buffer, decapping enzyme reaction buffer, siRNA reaction buffer, RT-PCR reaction buffer, or combinations thereof. Supportive agents may be combined into a single container or provided in separate containers. Preferably, the active agents are provided in separate containers.

[0058] In another embodiment, containers may contain siRNAs or sources for producing siRNA. The siRNA may be species specific. Any siRNA known in the art or yet to be discovered may be provided with the kit.

[0059] In another embodiment, containers may contain total RNA for control RT-PCRs to assess lariat purification. The total RNA may be from any species.

[0060] In another embodiment, containers may contain oligonucleotides, or primers, for control RT-PCRs. Such primers will amplify a well characterized linear RNA, lariat RNA, or combinations thereof, depending upon the control desired. One skilled in the art will recognize that the primers may be species specific and may depend upon the source species of the total RNA. For example, if the source of the total RNA is Saccharomyces cerevisiae, then the control primers could be those that would amplify ACT1 mRNA and the ACT1 intron lariat RNA.

[0061] The article of manufacture may also contain instructions of use.

DEFINITIONS

[0062] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art. All patents, applications, published applications and other publications are incorporated by reference in their entirety. In the event that there is a plurality of definitions for a term herein, those in this section prevail unless stated otherwise.

[0063] As used herein, the phrase “metallophosphatase domain” refers to the amino acids that are conserved among debranching enzymes isolated from various species.

[0064] As used herein, the term “enrich” or forms thereof refer to increasing the amount of a substance found in a heterogeneous population. For example, enriching for lariat RNA in an RNA population refers to increasing the proportion of lariat RNA in an RNA population to a proportion above the other types of RNA found in the RNA population. The enrichment includes purifying an RNA population to only include a specific type of RNA, such as lariat RNA.

[0065] As used herein, the term “hybridizes under stringent conditions” is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A, non-limiting example of stringent hybridization conditions are hybridization in 6x sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2x SSC, 0.1% SDS at 50-65°C (e.g., 50°C or 60°C or 65°C). Preferably, the isolated nucleic acid molecules of the invention that hybridizes under stringent conditions corresponds to a naturally-occurring nucleic acid molecule. As used herein, a naturally-occurring nucleic acid molecule refers to RNA or DNA molecules having a nucleotide sequence that occurs in a human cell in nature (e.g., encodes a natural protein).

[0066] As used herein, the phrase “lariat RNA” refers to the pre-mRNA that is excised during the formation of mRNA. This excised pre-mRNA forms a lariat structure.

[0067] As used herein, the phrase “linear RNA” refers to RNA that does not form a lariat structure and that can be degraded by exonucleases.

[0068] As used herein, the phrase “linear RNA degrading enzyme” refers to any enzyme capable of degrading linear RNA. Such enzymes include, without limitation, 3’ exonucleases, 5’ exonucleases, exonucleases with 3’ and 5’ activity, as well as others known in the art or yet to be discovered.

[0069] As used herein, the term “nucleic acid sequence” is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA or lariat) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded.

[0070] As used herein, the phrase “RNA population” refers to a sample containing ribonucleic acid. The RNA population may or may not be purified RNA.

[0071] The following example are simply intended to further illustrate and explain the present invention. The invention, therefore, should not be limited to any of the details in these examples.

EXAMPLES

Example 1

Materials and Methods

[0072] Yeast and Bacterial Strains, Plasmids, and General Procedures.

[0073] The following yeast strains were used: TMY30 (MATa ura3-52 lys2-801 ade2-101 trpl-D63 his 3-Δ 200 leu2-Δ1), TMY60 (TMY30 dbr1::neo), TMY497 [TMY30 mutated to dbr1 (D180Y allele)], TMY498 [TMY30 mutated to dbr1 (G84A allele)], TMY499 [TMY30 mutated to dbr1 (Y08S allele)]. TMY453, a dbr1a::hisG version of sigma
strain 10560-23C, was used for FLO8 RT-PCR experiments (sigma strain 10560-23C=MAlpha ura3-52 his3::hisG leu2::hisG). The drbr1Δ:hisG allele was created using pTM513, a DBR1 gene blastic plasmid containing drbr1 Δ:hisG-URA3-hisG, and targeted to replace DBR1 chromosomal sequences by digestion with PvuII.

[0074] The following E. coli strains were used: Rosetta DE3 (F ompT hsdSB (rB mB) gal dcm (DE3) pLyS SRAE (ComP); XL1-Blue (F-:Tn10 proA rB lacIq Δ[lacZΔM15 recA1 endA1 gyrA96 (NaI) thi hsdR17 (rK mB) supE44 relA1 lacI) JM109 F- tri dN36 lacYq Δ[lacZΔM15 proA rB lacIq14 (McrA) Δ[lac-pro-AB]) thy1A96 (NaI) endA1 hsdR17 (rK mB) relA1 supE44 recA1]) E. coli strain BL21-Codon Plus(DE3) RIL. Sonication of cells was performed on ice for 60 sec., in 1 sec. pulses, with a large probe at 50% power. Triton X-100 was added after sonication to a final concentration of 0.1%. The tagged Dbr1p was purified from E. coli extracts by binding to and eluting from Nickel-nitrotriacetic acid-agarose columns. Fractions were assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Peak fractions from the elution were dialyzed against debranching buffer (20 mM HEPES KOH, pH 7.9: 125 mM KCl: 0.5 mM MgCl2: 1 mM DT; 10% glycerol). In some cases, Dbr1p was concentrated by spinning through a Microcon YM-30 spin concentrator at 14,000×g for 40 min. at 4°C. In a Beckman Allegra 25R centrifuge (TA-15.1.5 rotor). The concentrations of Dbr1p preparations were 50-100 ng/μl. Mass spectrometry of purified Dbr1p was performed.

[0081] Enzymatic Treatments of RNA.

[0082] Bacillus steatorrhophilus PNPase was acquired (Sigma, St. Louis, Mo.) and a stock of 3.5 units/ml was prepared by dissolving the protein in water, then adding Tris HCl, pH 8.5, to a final concentration of 50 mM. PNPase reactions were performed in PNPase buffer (50 mM Tris HCl, pH 8.5; 1 mM 2-mercaptoethanol; 1 mM EDTA; 20 mM KCl: 15 mM MgCl2: 10 mM Na2 HPO4, pH 8.3) on 20-1000 ng of total yeast RNA in 20 μl reactions for 4-5 hr at 60°C, using 1 μl of the PNPase stock. Upon completion of reactions, samples were subjected to 85°C. for 10 min, then either used directly in RT-PCRs or ethanol precipitated. Mock treatments were performed in the same way, minus PNPase.

[0083] Approximately 50-100 ng of yeast Dbr1p prepared from E. coli was used for in vitro debranching reactions of 20-200 ng of RNA. Reactions were performed at 30°C. for 45 min. in a 20 μl volume containing lx debranching buffer (20 mM HEPES-KOH pH 7.9, 125 mM KCl, 0.5 mM MgCl2, 1 mM DT; and 10% glycerol). Reactions were stopped by heating at 65°C. for 10 minutes (min.), Mock treatments were performed in the same way, minus Dbr1p.

[0084] For sequential enzymatic treatments, RNA samples were phenol/chloroform extracted and ethanol precipitated after the first treatment (PNPase or Dbr1p) then resuspended and treated with the second enzyme.

[0085] RT-PCR Methods.

[0086] RT-PCRs of Lariat and linear RNAs were performed using QIAGEN one-step RT-PCR kit (Valencia, Calif.) under the following general conditions: 50°C, 30 min; 95°C, 15 min; 9 cycles of 94°C. for 30 sec, 54°C. for 30-60 sec (touchdown to 46°C. (-1°C. per cycle)), 72°C. for 30 sec; X cycles (see below) of 94°C. for 30 sec, 46°C. for 30 sec, 72°C. for 3045 sec; 72°C. for 5-10 min; 4°C. hold. The number of cycles in the post-touchdown phase of different RT-PCRs (X cycles above) varied with the experiment and are reflected in the following reaction profile names: ACTI-1, 29 cycles, post-touchdown; ACTI-2, 24 cycles, post-touchdown; ACTI-3, 19 cycles, post-touchdown; ACTI-4, 15 cycles, post-touchdown; and ACTI-5, 11 cycles, post-touchdown. RNA amounts between 2 ng and 50 ng were used in RT-PCRs. RT-PCRs were analyzed by either PAGE or agarose gel electrophoresis.

[0087] Real-Time RT-PCR (qRT-PCR) of Lariat and Linear RNAs.

[0088] Primers and probes for qPCR were designed using Sequence Detection Systems software from Applied Biosystems (Carlsbad, Calif.) and are listed in Table 1. All probes and primers for qRT-PCR were purchased from Applied Biosystems. Validation experiments were performed that demonstrated that the efficiencies of target and reference PCRs are approximately equal.
TABLE 1. Primers and probes for qRT-PCR.

<table>
<thead>
<tr>
<th>Target SEO and ID</th>
<th>Primers NO:</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 FWD primer</td>
<td>TCCCAAGATCGAATAATTACTGAA</td>
<td>-30 to 6</td>
<td></td>
</tr>
<tr>
<td>29 REV primer</td>
<td>TTTACCATACAGAGGGCCCTAAT</td>
<td>54 to 28</td>
<td></td>
</tr>
<tr>
<td>30 TaqMan probe</td>
<td>VIC-TGAATTAACAGAGGGCCCTGCTC-MGB</td>
<td>-4 to 26</td>
<td></td>
</tr>
<tr>
<td>ACT1 intron</td>
<td>AAGCCCTGGTGGTGTAATG</td>
<td>89 to 106</td>
<td></td>
</tr>
<tr>
<td>31 FWD primer</td>
<td>ATTTCTCAGTCTTCCAACTCCCTTATA</td>
<td>94 to 121</td>
<td></td>
</tr>
<tr>
<td>32 REV primer</td>
<td>TTTACAGCCGTATTATACTCAAT</td>
<td>173 to 150</td>
<td></td>
</tr>
<tr>
<td>33 TaqMan probe</td>
<td>6FAM-TGACAGTTAGCTGTTAAACC-MGB</td>
<td>123 to 142</td>
<td></td>
</tr>
<tr>
<td>RPP1B mRNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34 FWD primer</td>
<td>AGGCCCTGGTGGTGTAATG</td>
<td>89 to 106</td>
<td></td>
</tr>
<tr>
<td>35 REV primer</td>
<td>TCCAAAGCCTTACGTAACATC</td>
<td>146 to 124</td>
<td></td>
</tr>
<tr>
<td>36 TaqMan probe</td>
<td>VIC-CGACAGCGTCCTGCTG-MGB</td>
<td>108 to 122</td>
<td></td>
</tr>
<tr>
<td>RPP1B intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 FWD primer</td>
<td>AAAGCAACCTAAACAGACTTTTG</td>
<td>12 to 35</td>
<td></td>
</tr>
<tr>
<td>38 REV primer</td>
<td>TTCTCGGGACAGTTGTTGCT</td>
<td>77 to 57</td>
<td></td>
</tr>
<tr>
<td>39 TaqMan probe</td>
<td>6FAM-ACTACAGAGGAAGATT-MGB</td>
<td>38 to 55</td>
<td></td>
</tr>
<tr>
<td>YRA1 mRNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 FWD primer</td>
<td>AGTTTTGCAGAAGGCAGCTTAAAG</td>
<td>249 to 270</td>
<td></td>
</tr>
<tr>
<td>41 REV primer</td>
<td>ACCACCACTACCTAGATGAAGAAA</td>
<td>314 to 291</td>
<td></td>
</tr>
<tr>
<td>42 TaqMan probe</td>
<td>VIC-AGGATGGCTCTGAAAGGAAT-MGB</td>
<td>272 to 289</td>
<td></td>
</tr>
<tr>
<td>YRA1 intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43 FWD primer</td>
<td>GGCATGCTGCTGCTGGAT</td>
<td>42 to 60</td>
<td></td>
</tr>
<tr>
<td>44 REV primer</td>
<td>GATCAAGAGCCTGCGCATAC</td>
<td>107 to 86</td>
<td></td>
</tr>
<tr>
<td>45 TaqMan probe</td>
<td>6FAM-CAAGAAAATATGTTGAGAA-MGB</td>
<td>62 to 94</td>
<td></td>
</tr>
</tbody>
</table>

*Relative to start of coding sequence for mRNA primers and probes, relative to start of intron sequence for intron primers and probes.

[0090] PCR MasterMix reagents from Applied Biosystems were used for qPCR reactions, which were performed in triplicate for each sample. Reactions were prepared and run according to a standard protocol established by Applied Biosystems on an ABI 7500 real-time PCR machine. Briefly, reactions contained 2×PCR MasterMix, 900 mM forward primer, 900 nM reverse primer, 250 nM TaqMan probe, and cDNA (~20 ng). Reactions were incubated for 2 minutes at 50°C and then 10 minutes at 95°C and before proceeding through 40 cycles of a 30 second (sec) incubation at 95°C and a 60 second incubation at 60°C. Completed reactions were held at 4°C.

[0091] Relative quantification (RQ) of results was performed using the comparative CT method (ΔΔCT) (Schmittgen and Livak 2008). The amplification of each target intron sequence was compared to amplification of the corresponding mRNA sequence and a ΔCT was determined. To compare the different samples to each other, the wild-type sample was used as the calibrator sample. Therefore, the ΔCT of the wild-type sample was subtracted from the ΔCT for each sample to determine −ΔΔCT values. In FIG. 7, RQ 2−ΔΔCT for each −ΔΔCT represents the fold-difference in intron levels between a given sample and the wild-type sample (DBR1).

[0092] In Vitro Debranching Time Course.

[0093] Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA, the exogenous control for qPCR in these experiments, was generated by reverse transcribing 600 ng of human RNA at 45°C for 1 hour (hr) using the reverse transcriptase (RT) kit from Applied Biosystems. A debranching reaction mix was set up on ice and contained 5600 ng of total RNA from TMY60 (dbrlp) cells, about 6 ng GAPDH cDNA, 140 μl of purified Dbrlp, and 350 μl 2× debranching buffer in a final volume of 700 μl. Seven 100 μl aliquots of this mix were distributed to 0.2 ml PCR tubes. The debranching reaction was directly inactivated in one tube (0 min reaction time) by raising the temperature to 95°C, followed by phenol/chloroform extraction and ethanol precipitation. The remaining six tubes were incubated at 30°C and individual reactions were stopped after 2.5 min, 5 min, 10 min, 15 min, 30 min, and 60 min. Reactions were stopped by raising the temperature to 95°C, followed by phenol/chloroform extraction and ethanol precipitation. RNAs were then treated with PNPase, as described above, to degrade intron lariats linearized by Dbrlp. Reverse transcription of the RNAs remaining from the different debranching reactions was performed using the RT kit from Applied Biosystems and random hexamer primers. qPCRs using these cDNAs were performed as described above, amplifying a volume of cDNA roughly corresponding to about 20 ng of starting total RNA, using primers and probes for yeast ACT1, YRA1, and RPP1B introns as well as human GAPDH. GAPDH cDNA was the exogenous control because it is insensitive to PNPase and remained at a constant level in each reaction.

[0094] Creation of dbrlp Point Mutant Strains.

[0095] Mutants were created using modifications of the delito perfetto method (Storici et al. 2001) and the site specific genomic (SSG) method (Gray et al. 2004). Initially, a dbrlp::URA3 strain was created to facilitate the introduction of point mutant alleles of dbrlp into the DBR1 locus. Yeast strain TMY490, containing a URA3-marked deletion of 1090
bp of the 1215 bp DBRI coding sequence (nts 71-1160 deleted), was constructed by transformation of TMY30 with a PCR fragment containing the URA3 gene from pRS306 flanked by ends corresponding to 5' and 3' segments of the DBRI coding region.

[0096] The fragment used for making the dbrlΔ::URA3 allele was created by PCR of pRS306 with oligonucleotides 443 and 444, the 3' 20 nucleotide (nt) of which anneal to the ends of the URA3 gene on pRS306 and the 5' 40 nt of which correspond to DBRI sequences (see Table 2).

<table>
<thead>
<tr>
<th>SEQ NO.</th>
<th>Primer</th>
<th>Sequence</th>
<th>Position*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>146</td>
<td>cactctccataacaccccta</td>
<td>ACTI intron nt 100-119</td>
</tr>
<tr>
<td>2</td>
<td>215</td>
<td>ctcaacaccgaagacacagaa</td>
<td>ACTI nt -128 to -107</td>
</tr>
<tr>
<td>3</td>
<td>216</td>
<td>tgtatacttgtgttcatct</td>
<td>ACTI nt 130 to 109</td>
</tr>
<tr>
<td>4</td>
<td>331</td>
<td>aggagtggctctcttagaa</td>
<td>-761 to -741 upstream of DBRI ORF</td>
</tr>
<tr>
<td>5</td>
<td>332</td>
<td>gaggatctctgataatgtgctgccatct</td>
<td>-10 to -30 upstream of DBRI ORF; EcoRI site added at 5' end</td>
</tr>
<tr>
<td>6</td>
<td>333</td>
<td>gctctagaacagatggcaagcaggaatagga</td>
<td>16 to 30 after of DBRI stop codon; XbaI site added at 5' end</td>
</tr>
<tr>
<td>7</td>
<td>336</td>
<td>attaagaacagcagctccaccttag</td>
<td>779 to 760 after of DBRI stop codon; NotI site added at 5' end</td>
</tr>
<tr>
<td>8</td>
<td>363</td>
<td>gcaagcttagaacaactcttag</td>
<td>ACTI intron nt 16-1, 265-262</td>
</tr>
<tr>
<td>9</td>
<td>372</td>
<td>agtgaattcctgactccacattc</td>
<td>FLO8 nt 12-35</td>
</tr>
<tr>
<td>10</td>
<td>373</td>
<td>cataaaaaagctgtgagttg</td>
<td>FLO8 nt 418-398</td>
</tr>
<tr>
<td>11</td>
<td>374</td>
<td>ggtgcaaaatattctgtgcacatct</td>
<td>FLO8 nt 422-445</td>
</tr>
<tr>
<td>12</td>
<td>375</td>
<td>atcctggttggcctgctactt</td>
<td>FLO8 nt 837-816</td>
</tr>
<tr>
<td>13</td>
<td>376</td>
<td>aagtaaacagctactgtggtgg</td>
<td>FLO8 nt 841-861</td>
</tr>
<tr>
<td>14</td>
<td>377</td>
<td>tgtctgatccggagagttag</td>
<td>FLO8 nt 1260-1241</td>
</tr>
<tr>
<td>15</td>
<td>378</td>
<td>tggccgaggagatattatttc</td>
<td>FLO8 nt 1268-1289</td>
</tr>
<tr>
<td>16</td>
<td>379</td>
<td>aagtaaatggcgactggtagatcagcag</td>
<td>FLO8 nt 1675-1652</td>
</tr>
<tr>
<td>17</td>
<td>360</td>
<td>ttcatcggacagatggcAA</td>
<td>FLO8 nt 1693-1712</td>
</tr>
<tr>
<td>18</td>
<td>361</td>
<td>ttctcctggagttagatagag</td>
<td>FLO8 nt 2036-2013</td>
</tr>
<tr>
<td>19</td>
<td>362</td>
<td>atcaaggtatcgactttgacgc</td>
<td>FLO8 nt 2054-2075</td>
</tr>
<tr>
<td>20</td>
<td>363</td>
<td>cagccttccctaatatatgaaat</td>
<td>FLO8 nt 2399-2376</td>
</tr>
<tr>
<td>21</td>
<td>408</td>
<td>taataatgtcagcccagaacagg</td>
<td>URA3 nt 67-46</td>
</tr>
<tr>
<td>22</td>
<td>417</td>
<td>ttgcaatgtgtgtcacaagg</td>
<td>DBRI nt 10-29</td>
</tr>
<tr>
<td>23</td>
<td>418</td>
<td>caagtcgaattgagaattgtagatgc</td>
<td>DBRI nt 1217-1192</td>
</tr>
<tr>
<td>24</td>
<td>443</td>
<td>gctctgatgtgctgccagctcAA</td>
<td>5' 40 nt = DBRI nt 31-70</td>
</tr>
<tr>
<td>25</td>
<td>444</td>
<td>taactatggccagctcagagc</td>
<td>3' 20 nt = URA3 flank in pRS306</td>
</tr>
<tr>
<td>26</td>
<td>444</td>
<td>gataaatgttttagttgctgtacttactttctgata</td>
<td>5' 40 nt = DBRI nt 1200-1161</td>
</tr>
</tbody>
</table>
**TABLE 2 - continued**

**Oligonucleotides.**

<table>
<thead>
<tr>
<th>SEQ ID NO: Primer</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>27 444 cont ... cctgatggttcttct</td>
<td>3' 20 nt = URA3 flank in pRS306</td>
<td></td>
</tr>
</tbody>
</table>

*For the ACT1, FLO8, URA3 and DBR1 genes, the nucleotide positions are relative to the first nucleotide of the coding sequence, except for the ACT1 intron, where positions are relative to the first nucleotide of the intron.*

---

**[0097]** The **dbr1Δ::URA3** disruption on yeast chromosome XI was created by homologous recombination between the **DBR1** locus and the **dbr1Δ::URA3** PCR fragment. Briefly, TMY30 was transformed with the **dbr1Δ::URA3** PCR fragment and transformants were selected on SD-Uracil plates. Transformants were screened by PCR with primer pairs 401/402, which anneal within the **DBR1** sequences that are deleted in the **dbr1Δ::URA3** allele, and 417/418, which anneal outside the **DBR1** sequences that are deleted in the **dbr1Δ::URA3** allele. Transformants containing the **dbr1Δ::URA3** allele template a 417/418 PCR product but not a 401/402 PCR product. DNA sequencing of PCR products was performed to verify the presence of the **dbr1Δ::URA3** allele.

**[0098]** Replacement of the chromosomal **dbr1Δ::URA3** allele with **dbr1** point mutations was accomplished by transformation. TMY490 (**dbr1Δ::URA3** strain) was co-transformed with **YEp351** (LEU2) and PCR fragments of **dbr1** point mutants. The PCR fragments were generated from plasmids **pTM431**, **pTM432**, and **pTM435** with PCR primer pairs 417/418. Transformants (with **YEp351**) were selected in **SD-leucine** liquid media during a 48 hr incubation period at 30°C (w/ shaking). After this selection period, cells were spread onto 5-fluoroorotic acid plates to select for cells that lost function of the **URA3** gene within the **DBR1** locus. Recombinants within the **FOA** population that have replaced the **dbr1Δ::URA3** allele with a **dbr1** point mutant allele were identified by PCR screening. Positive clones were identified as those that template a 417/418 PCR product but not a 417/408 PCR product (specific for the **dbr1Δ::URA3** allele). DNA sequencing of PCR products was performed to verify the presence of a **dbr1** point mutant allele.

**Example 2**

**RT-PCR Detection of Lariat RNAs**

**[0099]** *Saccharomyces cerevisiae* **ACT1**, which encodes actin, is a robustly expressed gene that contains an intron of 308 nt. The first example of a spliceosomal intron discovered in yeast, the **ACT1** intron contains all the canonical features of yeast introns and is efficiently spliced from pre-mRNA, producing an excised lariat with a 265 nt circle. This well-characterized gene was chosen to assess intron levels as tools were developed and tested for detecting and enriching excised intron lariats. Primers were designed for use in RT-PCR to detect the lariat form of the **ACT1** intron RNA and, as a control, **ACT1** mRNA (FIG. 1A). RT-PCR of total yeast RNA using primers that flank the **ACT1** exon-exon junction (primers 215 and 216) amplifies a 285 bp product from **ACT1** mRNA. Primer 363 spans the **ACT1** intron lariat branch point and is used in combination with primer 146, which anneals to sequences complementary to the **ACT1** intron upstream of the lariat branch point, in an RT-PCR that amplifies a 184 bp product from the lariat form of the **ACT1** intron RNA. As expected, when RT-PCR reactions were performed using total RNA samples from wild-type (TMY30) and **dbr1** mutant yeast cells (TMY60), the amounts of **ACT1** mRNA products are similar when using equivalent amounts of RNA from the two cell types (FIG. 1B, lanes 1 and 3 as well as lanes 5 and 7). However, the **ACT1** intron lariat product is much more readily produced from **dbr1** cells (FIG. 1B, lane 4 vs. 2 and lane 8 vs. 6). These data clearly show that a **dbr1** mutant strain or, where appropriate, a **dbr1**p knock-down strain contains a rich source of expressed intron sequences. It is also evident that the use of intron-specific RT-PCR could be used to detect excised introns from genes expressed at very low levels. For studies on alternative splicing, the use of RT-PCR on RNA from **dbr1**p-deficient cells can allow detection of rare splice variants.

**[0100]** A previous report described the use of radiolabeled primers spanning intron RNA branch points for analyzing intron populations by primer extension (Spingola et al. 1999). The RT-PCR method we describe could be modified to survey intron lariats containing specific sequences at intron 5' ends and branch points. RT-PCR has added utility because the products can be cloned and sequenced to identify the individual introns represented in a lariat population.

**Example 3**

**Insensitivity of Lariat RNAs to the 3’ Exonuclease PNPase**

**[0101]** Linear and lariat RNAs have different sensitivities to 3' exonucleases, including PNPase, a component of bacterial RNA degradation systems. PNPase degrades linear RNAs but does not proceed past the 2' branch present in intron RNA lariats. Therefore, treatment of RNA samples with an enzyme like PNPase should result in a vast enrichment of excised intron lariats in the RNA that remains intact after treatment. This difference should be evident in the results of the RT-PCR assay described above when amplifying PNPase-treated RNA samples. Since RNA secondary structures reduce the efficiency of PNPases, reactions were performed at elevated temperature (60°C) using PNPase from *Bacillus stearothermophilus* to circumvent this problem. Total RNA samples from a **dbr1** mutant strain (TMY60) were treated with a range of PNPase concentrations and then subjected to RT-PCR to detect **ACT1** intron lariat RNAs as well as the linear mRNA (FIG. 2). Results are consistent with expectations that the use of PNPase selectively preserves RNA lariats.

**[0102]** The high temperature reaction using PNPase from a thermophile appears to be much more efficient than the reported reaction with the *E. coli* PNPase at 37°C. In order to
eliminate the RT-PCR product from the ACT1 mRNA, PNPtr must degrade, at the very least, the RNA corresponding to the binding site for the downstream primer (oligonucleotide 216). To accomplish this, PNPtr must degrade all the RNA that lies to the 3' side of the oligonucleotide 216 binding site, which includes 998 nt of the ACT1 coding sequence plus the 3' UTR and the polyA tail. To further examine the processivity of *Bacillus stearothermophilus* PNPtrase, the degradation of FLO8 mRNA was assessed. FLO8 mRNA is 2.4 kb in length. Primer pairs were designed to amplify different portions of this mRNA along its length (Fig. 3A). Total nucleic acid samples and RNA samples (DNased total nucleic acid samples) were treated with PNPtrase and subjected to RT-PCR to detect the various segments of FLO8. As shown in Fig. 3B, PNPtrase readily degrades every segment of FLO8 mRNA assayed. As expected, PNPtrase has no effect on FLO8 DNA present in the total nucleic acid samples (Fig. 3C). Other enzymes that worked as well as *Bacillus stearothermophilus* PNPtrase in our studies are *Thermus thermophilus* PNPtrase at 65°C and *Escherichia coli* RNase R at 37°C.

Example 4

Sensitivity of Lariat RNAs to Dbr1p

In order to perform Dbr1p treatments, *S. cerevisiae* Dbr1p was expressed in *E. coli* and purified by metal affinity chromatography (Figs. 4A and 4B). Although histidine-tagged Dbr1p is expected to have a mass of about 50 kilodalton (kDa), the mobility of the main product in SDSPAGE is about 45 kDa. Others have observed this anomalous mobility for histidine-tagged Dbr1p and have speculated that the protein may undergo limited proteolysis in *E. coli*. However, mass spectrometric analysis of the main band in the stained gel shows it to be the expected molecular mass of the histidine-tagged Dbr1p (50,062 Dalton (Da)) (Fig. 4C), indicating that the protein is intact and must run anomalously in SDS-PAGE because of its physical properties.

Using the Dbr1p enzyme preparation, debranching reactions were carried out on total RNA samples from a dbr1 mutant strain. RT-PCR analysis reflects the differential sensitivity of linear and lariat RNAs to Dbr1p. After Dbr1p treatment, RT-PCR detection of ACT1 RNA lariat is greatly decreased (Fig. 5, lane 4 vs. 2). On the other hand, the product indicative of ACT1 linear mRNA is still readily detectable after Dbr1p treatment (Fig. 5, lane 3 vs. 1).

Example 5

Combinations of PNPtrase and Dbr1p Treatments

PNPtrase and Dbr1p treatments can be used in combination when exploring the properties of a particular RNA species. Sequential enzymatic treatments can also be used to enrich for RNA lariats and then linearize them for further manipulations. To demonstrate this, ACT1 RNA species present within a total RNA sample from a dbr1 mutant strain were analyzed by RT-PCR following sequential PNPtrase and Dbr1p treatments. As shown in Fig. 6A (lanes 1-4), initial treatment of the RNA sample with PNPtrase degrades the linear mRNA (lanes 1 and 3), but leaves lariat RNA intact (lane 2). Subsequent treatment with Dbr1p shows that the resistant RNA is a lariat (lane 4). As shown in Fig. 6A (lanes 5-8), skipping the initial PNPtrase treatment leaves the linear mRNA intact (lanes 5 and 7) as well as the lariat RNA (lane 6). The lariat RNA is then distinguished by its sensitivity to cutting with Dbr1p (lane 8). The order of the PNPtrase and Dbr1p reactions can be switched to generate a complementary set of predictable results (Fig. 6B).

Example 6

Real-Time RT-PCR Measurement of Lariat RNA Levels

A real-time RT-PCR method (qRT-PCR), using the TaqMan detection system (Applied Biosystems), was developed to quantitatively compare the intron RNA lariat levels of different samples. The study included not only the ACT1 intron but also the YRA1 and RPPIB introns to investigate the generality of the methods. YRA1 encodes an RNA binding protein involved in mRNA export from the nucleus and is moderately expressed, although less than ACT1. The YRA1 intron is 765 nt in length, which is larger than the 300 nt average for yeast introns, and contains a non-canonical branch point sequence. Furthermore, the intron is inefficiently spliced from pre-mRNA, which is important for the auto-regulation of Yra1p protein levels. RPPIB encodes a ribosomal protein and is even more highly expressed than ACT1. The RPPIB intron is typical for yeast, 301 nt in length, with canonical sequences.

Initially, a strategy similar to the one used for RT-PCR of ACT1 intron lariats described above, with one primer spanning the lariat branch point and serving as both the RT primer and the reverse primer for PCR was used. However, a different strategy using random primers for the RT step was also used to allow amplification of the different target sequences from a common pool of cDNA. Consequently, both PCR primers anneal upstream of the branch point for each target gene, with a TaqMan probe annealing between them (Fig. 7A). Since these types of primers will also prime amplification of genomic DNA we run control PCRs for each sample without a prior RT step to ensure that DNA contamination was not contributing to the PCR product. The mRNA for each target gene served as the endogenous control for qRT-PCR (Fig. 7A, top). Using this strategy, intron sequences for ACT1, RPPIB, and YRA1 were amplified from dbr1 and wild-type yeast strains (TMY60 and TMY30). As shown in Figs. 7B, 7C and 7D [Dbr1 (wild type) vs. dbr1 null mutant], the real-time method generated the expected results: the different intron RNAs accumulate at higher levels in the dbr1 null mutant strain than in wild type.

Recently, qRT-PCR was also used to analyze mutant variants of Dbr1p. Previously, a set of point mutants had been created by random PCR mutagenesis and analyzed for intron RNA levels by an RNase protection assay. In these experiments, the dbr1 mutant alleles were under the control of a strong, inducible promoter (pGAL1) and carried on a high copy plasmid.
The yeast strain carried a dbr1Δ mutation [open reading frame (ORF) deletion] at the DBR1 locus so the plasmid-borne dbr1 mutant alleles were the only sources of Dbr1p. For the current study, three dbr1 point mutants (D180Y, G84A, and Y68S) were analyzed by qRT-PCR to determine their levels of intron lariat RNA relative to wild-type (DBR1) and dbr1Δ. To make the analysis more biologically relevant, each of the dbr1 mutant alleles was placed at the DBR1 locus, replacing the wild-type allele, and was under the control of the native DBR1 promoter. After log-phase growth of cells, RNA samples from wild-type and mutant strains were harvested and subjected to qRT-PCR to amplify intron and messenger RNA sequences from ACT1, RPP1B, and YRA1. The three dbr1 alleles tested show strong intron RNA accumulation phenotypes, comparable to the dbr1Δ knockout allele (Figs. 7B, 7C, and 7D).

Example 7

qRT-PCR Analysis of a Debranching Time Course

[0110] Using a combination of Dbr1p and PNPase treatments, in vitro debranching reactions of total cellular RNA from a dbr1 strain were followed over time courses of thirty minutes. Debranching reactions were stopped at different times and the reaction products were treated with PNPase to degrade linearized intron RNAs. The remaining intron lariats were detected by qRT-PCR as described herein. Because the PNPase treatment step degrades all linear RNAs, human GAPDH cDNA was added to the yeast RNA samples as an exogenous control. The GAPDH cDNA is insensitive to both Dbr1p and PNPase, remaining at the same level in the various samples. Debranching of the ACT1 and RPP1B intron lariats was almost complete within the first 5 minutes of the reactions (Fig. 8). However, the debranching rate of the ACT1 intron lariat appeared to be only two-thirds the initial rate of the RPP1B intron lariat.

[0111] The results observed from using qRT-PCR to follow in vitro debranching, show that the debranching rates can vary from one intron lariat to another. The ACT1 intron is debranched at only two-thirds the initial rate at which the RPP1B intron lariat is debranched. These data suggest that different intron lariats are debranched at different rates in vivo, which may be of functional significance. Slower rates of debranching may occur for introns that contain snRNAs or mirtrons, reflecting the binding of additional factors to intron sequences or specific folding properties of the RNA. Thus, the rate of the debranching of introns can be used to predict which introns may contain additional information. Relative debranching rates can be inferred from quantitative analysis of intron RNA levels relative to mature mRNA levels for a given gene compared to a standard, rapidly debranching intron RNA. For these types of experiments, RNA samples could be taken from a wild-type strain (DBR1), where lariat RNAs are not stabilized. Inefficient splicing would have to be ruled out before further study of candidate slow debranchers. As described above, YRA1 is an example of a gene that uses splicing inefficiency to regulate protein levels.

[0112] Quantitative RT-PCR of lariat RNAs can be used to determine the relative rates of transcription for different intron-containing genes. The use of intron RNA levels as a novel data source for estimating relative levels of transcription for pre-mRNAs limits the utility to intron-containing genes, a notable limitation for S. cerevisiae. Furthermore, a Dbr1p-deficient strain would have to be used for intron lariats to be a stable record of transcription. Work with yeast dbr1 mutants over the years has not found any significant perturbation of cellular physiology other than the accumulation of intron RNA lariats. In the experiments shown in FIG. 7B-D, the level of RPP1B intron RNA in a dbr1 strain relative to the level in wild type is much greater (about 330-fold) than the corresponding levels of ACT1 and YRA1 intron RNAs (about 13-fold). These data indicate that the transcription rate for RPP1B is almost 30-fold greater than the rates for ACT1 and YRA1 (summarized in Table 3). These relative transcription rates are very different from estimates based on nuclear run-on assays, mRNA steady state levels plus half-lives, and DTA (Table 3).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Transcriptional frequency</th>
<th>DTA</th>
<th>Relative intron levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT1</td>
<td>45.5d (1)</td>
<td>7.2e (1)</td>
<td>63.2 (1)</td>
</tr>
<tr>
<td>YEL039C</td>
<td>16.2d (0.4)</td>
<td>80.6f (11.2)</td>
<td>88.9 (1.4)</td>
</tr>
<tr>
<td>YDR388W</td>
<td>120.0d (2.6)</td>
<td>23.0f (3.2)</td>
<td>192.7 (3)</td>
</tr>
</tbody>
</table>

*In RNA/cell/hr; numbers in parentheses are levels normalized to ACT1 level.

*DTA = dynamic transcription analysis, measured as mRNAs/cell/cycle time (150 min); numbers in parentheses are levels normalized to ACT1 level.

*Derived from data in FIG. 7 for the dbr1 null strain versus wild type for each gene and normalized to ACT1 level.

*Estimated from RNA expression levels and mRNA half-lives.

*Estimated from genomic run-on experiments.

[0113] An area where the utility of excised introns is clearer is in determining relative rates of alternative splicing for a particular gene. Variable stabilities of different mRNAs confound estimates of their rate of synthesis, whether the synthesis that produces the mRNAs in question is transcription or alternative splicing. The use of a Dbr1p-deficient strain, which stabilizes the alternatively excised intron lariats equivalently, results in intron RNA lariat levels that directly reflect the rate of alternative splicing.

[0114] The methods described herein can also be applied to genome-wide analysis of introns themselves and are an improvement on previous analyses that also directly analyzed intron RNA lariats. RNA-seq of intron RNA lariat populations prepared using PNPase can provide complementary information to RNA-seq of whole transcriptomes and may reveal new lariat sequences not evident from transcriptome analysis alone. Intron RNA lariat levels can be greatly enhanced by blocking the RNA debranching reaction, which increases the likelihood of detecting even rare splicing events. Because cells defective for RNA debranching activity accumulate excised introns in their lariat forms, with shorted 3' tails, information on the 3' intron-exon junction is not obtainable from intron RNA lariat sequences. Nevertheless, lariat sequences provide information about branch points that is not obtainable from whole transcriptome sequencing. Such information is especially useful for studies of introns in organisms whose branch point sequences are not as highly conserved as those in S. cerevisiae [e.g. humans]. Finally, the absence of known intron sequences from an RNA population enriched-for RNA lariats can indicate that a gene is not expressed under the growth regimen employed. However, if an intron-containing gene is known to be expressed during the experiment, absence of intron sequences from the RNA lariat population...
could be an indication that the intron is removed by the hydrolytic splicing pathway observed for self splicing group II introns rather than the predominant branching pathway. High-throughput sequencing of enriched lariat RNAs from human cells is useful for much more detailed analysis of human branch point sequences.

Example 8

Amino Acid Conservation Among RNA Debranching Enzymes

[0115] Dburl is an RNA lariat debranching enzyme that hydrolyzes 2'-5' phosphodiester bonds at the branch points of excised intron lariats. The alignment model shown in FIG. 14 represents the N-terminal metallophosphatase domain of Dburl. This domain belongs to the metallophosphatase (MPP) superfamily. MPPs are functionally diverse, but all share a conserved domain with an active site consisting of two metal ions (usually manganese, iron, or zinc) coordinated with octahedral geometry by a cage of histidine, aspartate, and asparagine residues. The MPP superfamily includes: Mre11/SbcD-like exonucleases, Dburl-like RNA lariat debranching enzymes, Ycfl-like phosphodiesterases, purple acid phosphatases (PAPs), Ybbf-like UDP-2,3-diacetylglucosamine hydrolases, and acid sphingomyelinases (ASMases). The conserved domain is a double beta-sheet sandwich with a di-metal active site made up of residues located at the C-terminal side of the sheets. This domain allows for productive metal coordination.

Example 9

Creation of an RNA Sample that is Highly Enriched for RNA Lariats

[0116] Linear and lariat RNAs have different sensitivities to exonucleolytic enzymes. Almost all linear RNAs are susceptible to complete or nearly complete degradation by enzymes that have 3' exonuclease activity. The key is to use an enzyme that is blocked by the lariat branch point and cannot degrade past the branch point. Such enzymes leave the lariat loop intact. Since lariat RNAs lack a 5' end, they are insensitive to 5' exonuclease activity; however, many linear RNAs are also sensitive to 5' exonuclease activity. Linear RNAs with 5' cap structures, which protect RNAs from 5' exonuclease activity, are made sensitive to 5' exonuclease activity by removing their caps. Cap removal treatments do not make lariat RNAs sensitive to exonucleases with 5' exonuclease activity. FIG. 9 and FIG. 10 illustrate the 3' and 5' exonuclease strategies, respectively, to create an RNA population that is highly enriched for RNA lariats.

[0117] Several methods may be employed to create an RNA population that is highly enriched for RNA lariats. Such methods include: 1) treat the RNA sample with a nuclease that has 3' exonuclease activity, a combination of nucleases with 3' exonuclease activity can also be used; 2) treat the RNA sample to remove the 5' cap structure from mRNAs, then treat with a nuclease that has 5' exonuclease activity, a combination of nucleases with 5' exonuclease activity can also be used; 3) treat the RNA sample with a nuclease that has both 3' and 5' exonuclease activity, with or without prior treatment to remove the 5' cap structure from mRNAs, a combination of nucleases, one or more with 5' exonuclease activity plus one or more with 3' exonuclease activity, can also be used.

[0118] To increase the proportion of lariat RNAs in the RNA population, RNA samples can be obtained from cells in which RNA debranching enzyme activity has been lowered or eliminated. Because the enhancement of RNA lariat levels in these cells is so dramatic, the resulting sample is useful for identifying RNA lariat species that are normally present at very low levels (i.e. in cells that have wild-type levels of RNA debranching enzyme activity). Another method that can be employed to increase the proportion of lariat RNAs in the RNA population being studied is to selectively remove RNA species from the RNA sample prior to lariat RNA enrichment.

[0119] A control RNA sample that is depleted of lariat RNAs can be created and processed in parallel to the exonuclease-treated RNA sample to identify which RNAs are lariats in the exonuclease-resistant RNA population. The lariat depleted RNA sample is created by treatment of an RNA sample with RNA debranching enzyme prior to the exonuclease treatment protocol.

[0120] Following acquisition and treatment of RNA samples as outlined above, RNAs are processed for high-throughput sequencing. Although different platforms for high-throughput sequencing have been developed and continue to be developed, all of platforms involve parallel sequencing of large numbers of DNA fragments. All of these platforms are used for RNA sequencing by incorporating cDNA production protocols. The lariat-seq technique requires conversion of lariat-enriched RNA samples into cDNA populations, which are then processed for high-throughput sequencing according to the methods developed for the individual high-throughput sequencing platforms.

[0121] To aid in determining which cDNA sequences represent lariat RNAs, sequencing data resulting from experimental and control RNA samples are compared. RNAs originally in a lariat conformation will be represented at lower levels (proportionally and absolutely) in the control sample, resulting in a proportional (and absolute) reduction in the number of their corresponding cDNA sequences relative to the experimental sample. Some RNAs that are not in a lariat conformation in samples extracted from cells will survive the treatments to create an RNA population enriched for RNA lariats, for example RNAs with covalent modifications, other than a 2'-5' branch, that block the exonuclease used to create the experimental sample. These RNAs will be represented approximately equally (in absolute terms) in experimental and control samples. Selective removal of rRNA species from the RNA sample prior to lariat RNA enrichment, as stated above, will remove many RNAs that contain non-lariat covalent modifications that block the exonuclease.

[0122] Further evidence that a nucleic acid-resistant RNA identified by lariat-seq has a lariat conformation comes from signature cDNA products unique to lariat RNAs. Reverse transcriptase (RT) used for creating cDNA for sequencing is blocked by the presence of a 2' branch in an RNA substrate. However, when traveling along the branch segment itself, RT will read across the 2'-5' bond, creating cDNAs that juxtapose sequences that are not contiguous in the reference genome. Furthermore, when RT reads across the 2'-5' bond it inserts a nucleotide that is not expected according to Watson-Crick base pairing rules. Typically, for an intron lariat branch point, RT inserts an A opposite the branch point A instead of a T. Sequence reads that contain discontinuous genome segments with an unexpected nucleotide at the junction of the two segments are evidence that the cDNA was created from a lariat RNA.
The different sensitivities of linear and lariat RNAs to the 3' exonuclease polyribonucleotide phosphorylase (PNPase) are shown in FIG. 11. Using RT-PCR to measure RNA levels, it is apparent that exonuclease treatment degrades a linear RNA down to the limit of detection while a lariat RNA remains virtually untouched (comparing lanes 1 and 2 (PNPase treatment) to lanes 5 and 6 (untreated)).

Linear and lariat RNAs also have different sensitivities to RNA debranching enzyme, which can be exploited to confirm that an RNA species has a lariat conformation. In vitro cleavage of intron RNA lariats with purified S. cerevisiae Dbr1p is readily detectable with an RT-PCR assay, as shown in FIG. 12, lanes 3 and 4 (Dbr1p treatment) versus lanes 1 and 2 (untreated). For the RT-PCR in FIG. 12, a primer that spans the branch point was used for RT-PCR, which is why the intron signal is reduced upon Dbr1p treatment.

The use of sequential Dbr1p and PNPase treatments to explore the properties of a particular RNA species (control described above) is depicted in FIG. 11. As shown in FIG. 11, lanes 3 and 4 show the loss of a known lariat RNA when Dbr1p treatment precedes PNPase treatment. Compare to lanes 1 and 2 (PNPase treatment only) as well as lanes 5 and 6 (no treatment).

Data from high-throughput sequencing of cDNAs created from PNPase-treated RNA samples support the feasibility and openability of lariat-seq. Total S. cerevisiae RNA from a strain lacking a functioning RNA debranching enzyme was converted into cDNA after PNPase treatment and subjected to a high-throughput sequencing protocol (Illumina platform). An example of a small portion of the results obtained is depicted in FIG. 13, which shows the sequence reads that match to a segment of chromosome 6 (FIG. 13A). What is striking about the results is that the only sequence reads that map to this 20 kb segment of the S. cerevisiae genome are from cDNAs that represent the intron region of the ACT1 gene (FIG. 13B). Furthermore, all these reads map within the sequences corresponding to the lariat loop of the intron; none of the sequence reads represent the 45 by that lie within the intron downstream of the lariat branch point. The fact that no sequence reads mapped to the ACT1 coding region or any other gene in the 20 kb segment depicted in FIG. 13 indicates how efficiently the 3' exonuclease degraded the linear RNAs in the sample.

Example 10

**RNA Lariat Enrichment Kit**

The components necessary for RNA lariat enrichment can be provided in a kit for ease of use. An example of such a kit is described below. Variations of the kit are also contemplated.

Components of RNA lariat purification and analysis kit include the following: *Bacillus stearothermophilus* polyribonucleotide phosphorylase (BspPNPase); 2x BspPNPase reaction buffer: 100 mM Tris HCl pH 8.5; 2 mM 2-mercaptoethanol; 2 mM EDTA; 40 mM KCl; 3 mM MgCl2; 20 mM Na2HPO4 pH 8.3; *Saccharomyces cerevisiae* RNA debranching enzyme (ScDbr1); 10x ScDbr1 reaction buffer: 200 mM Hepes KOH (pH 7.9), 1.25 M KCl, 5 mM MgCl2, 10 mM dithiothreitol; siRNAs (or siRNA sources) targeting mRNA for RNA debranching enzyme (different siRNA resources are packaged, depending on the organism for which the kit is specified); *Saccharomyces cerevisiae* total RNA samples (from dbr1 mutant and wild-type cells) for control RT-PCRs to assess lariat purification; primers for control RT-PCRs to amplify ACT1 mRNA (linear RNA) and the ACT1 intron lariat RNA from *Saccharomyces cerevisiae* total RNA samples; and, primers for control RT-PCRs for the organism for which the kit is specified to amplify a known linear RNA and a known lariat RNA.

The kit also includes instructions of use. An example of such instructions includes the following:

1. Grow cells for RNA preparation. Two growth conditions can be used, one in which expression of endogenous RNA debranching enzyme is reduced, causing intron lariats to accumulate, and one in which the endogenous RNA debranching enzyme expression is unperturbed. If applicable, deploy the supplied siRNA resources to create cells with enhanced RNA lariat levels.

2. Harvest cells and purify total cellular RNA. Alternatively, store cells after harvesting for future RNA purification.

3. Treat 1 nanogram-10 micrograms of total RNA with 10 units of BspPNPase in 1× BspPNPase reaction buffer for 60 minutes at 60°C.

4. Incubate completed BspPNPase reactions at 85°C for 10 minutes to inactivate the enzyme.

5. Phenol/chloroform extract RNA samples and ethanol precipitate them.

6. BspPNPase-treated RNA samples can be used for RT-PCRs of specific target RNAs (e.g. known linear and lariat RNAs) or for creation of cDNA libraries for Lariat-seq.

The kit may also include a control sample that is not enriched for RNA lariats. A control RNA sample that reflects the total RNA sample purified from cells is created by performing the above procedure but without BspPNPase in step 3.

The kit may also include a control sample that contains debranched RNA lariats. True lariat RNAs present in the BspPNPase-resistant RNA population will be sensitive to BspPNPase in RNA samples pretreated with ScDbr1.

1. Treat 1 nanogram-10 micrograms of total RNA with 10 units of ScDbr1 in 1× ScDbr1 reaction buffer in a 20 microliter reaction volume for 45 minutes at 30°C.

2. Incubate completed ScDbr1 reactions at 65°C for 10 minutes to inactivate the enzyme.

3. Phenol/chloroform extract RNA samples and ethanol precipitate them. Resuspend RNAs in 1× BspPNPase reaction buffer.

4. Continue with BspPNPase treatment as described in steps 3-6 above (Procedure for creating purified RNA lariats).

In order to Confirm the enrichment of RNA lariats, control RT-PCRs for known linear and lariat RNAs are performed on treated RNA samples (both the samples enriched for RNA lariats and the control samples). Primers are provided for use with *Saccharomyces cerevisiae* RNA as well as for the organism for which the kit is specified.

Following treatment of RNA samples as outlined above (and confirmation of lariat-enrichment), RNAs are processed for high-throughput sequencing. The next step is to create a cDNA library from each treated RNA sample using procedures established for the high-throughput sequencing platform to be used (Illumina, SOLiD, etc). Materials for
creating cDNA libraries are available from several different manufacturers.

[0144] Bacillus stearothermophilus polymerase phosphorolase (BspPNAse) storage buffer: 50% glycerol, 50 mM Tris-HCl (pH 8.5), 100 mM NaCl, 0.1 mM EDTA, 0.1% Triton X-100 and 1 mM dithiothreitol.

[0145] Saccharomyces cerevisiae RNA debranching enzyme (ScDbr1) storage buffer: 50% glycerol, 20 mM HEPES KOH (pH 7.9), 125 mM KCl, 0.5 mM MgCl2, 1 mM dithiothreitol.

[0146] One unit of BspPNAse activity is defined as the amount of PNAse that forms 1 µmol of ADP per hour at 60°C by depolymerizing of Poly A.

[0147] One unit of ScDbr1 activity is defined as the amount of ScDbr1 that debranches 50% of the ACT1 intron present in

1 microgram of a total Saccharomyces cerevisiae RNA preparation (from mid-log phase cells) from a dbr1 mutant strain per hour at 30°C.

[0148] The invention illustratively disclosed herein suitably may be practiced in the absence of any element, which is not specifically disclosed herein. It is apparent to those skilled in the art, however, that many changes, variations, modifications, other uses, and applications to the method are possible, and also changes, variations, modifications, other uses, and applications which do not depart from the spirit and scope of the invention are deemed to be covered by the invention, which is limited only by the claims which follow.

[0149] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.
Continued...

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, -10 - -30 upstream of DBR1 ORF, EcoRI site added at 5' end

<400> SEQUENCE: 5
gagcatctg ataatgtct gcctctc

<210> SEQ ID NO 6
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, 16-30 after of DBR1 stop codon, XbaI site added at 5' end

<400> SEQUENCE: 6
gcctagac gaatgcagac ggaattga

<210> SEQ ID NO 7
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, 779-760 after of DBR1 stop codon, NotI site added at 5' end

<400> SEQUENCE: 7
ataagacgc ggcgcgcaag gcctcgcag ttgtga

<210> SEQ ID NO 8
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, ACT1 intron nt 19-1, 265-262

<400> SEQUENCE: 8
gcaagcgca gaacacattc ag

<210> SEQ ID NO 9
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, FLO8 nt 12-35

<400> SEQUENCE: 9
gtgaatagt cgtatatcag attc

<210> SEQ ID NO 10
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, FLO8 nt 418-396

<400> SEQUENCE: 10
cataaaaaa gcctggagt g

<210> SEQ ID NO 11
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, FLO8 nt 422-445

<400> SEQUENCE: 11

ggtgcaaat attctgggac atct

<210> SEQ ID NO 12
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, FLO8 nt 837-816

<400> SEQUENCE: 12

attctgggt ggcctcatac tt

<210> SEQ ID NO 13
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, FLO8 nt 841-861

<400> SEQUENCE: 13

agcacaacg ttaactggtg t

<210> SEQ ID NO 14
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer, FLO8 nt 1260-1241

<400> SEQUENCE: 14

tgcttgattg oggaagt tag

<210> SEQ ID NO 15
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primers, FLO8 nt 1268-1289

<400> SEQUENCE: 15

ttgccgagga agatatttat tc

<210> SEQ ID NO 16
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primers, FLO8 nt 1676-1652

<400> SEQUENCE: 16

agataatgg actggataca gcceg

<210> SEQ ID NO 17
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primers, FLO8 nt 1693-1712

<400> SEQUENCE: 17

ttcgatccec aagtgccaa
ttttctcttg gatggataa ttgt

atcaggtata tgttttgac gc

cagccctccc aataataaa attg

taxatagctt ggcagcaaca gg

ttgcgaaatt ctgtacaagg

taagctatga attagagat aatgc

-continued
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR Primers, 5' 40 nt = DBR1 nt 31-70

SEQ ID NO 25
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR Primers, 3' 20 nt = URA3 flank in pRS306

SEQUENCE: 25

tgctgtgcag gtcagctaaa ccaaatatat aaagaagtgt

SEQ ID NO 26
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR Primers, 5' 40 nt = DBR1 nt 1200-1161

SEQUENCE: 26

taaatgctgg gcatcagagg

gataaatgt ttagtttgtg gtaactcatc tttcgatgta

SEQ ID NO 27
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: qRT-PCR Primers, FWD Primer

SEQUENCE: 27

tcgatgggatttactcc

tccccagatc gaaatttacctg

SEQ ID NO 28
LENGTH: 25
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: qRT-PCR Primers, Reverse primer at position 94-28

SEQUENCE: 28

tttgatacata ccgagacctg ttaaatcag

SEQ ID NO 29
LENGTH: 27
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: qRT-PCR Probe, TaqMan probe at position -4 to

SEQUENCE: 29

SEQ ID NO 30
LENGTH: 21
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: qRT-PCR Probe, TaqMan probe at position -4 to
-continued

tgattaaca aggtgtgcgc t

<210> SEQ ID NO 31
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: qRT-PCR Primers, FWD Primer at position 94-121

<400> SEQUENCE: 31
atatttctt ctcctatacctct tactata

<210> SEQ ID NO 32
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: qRT-PCR Primers, Reverse primer at position 173-150

<400> SEQUENCE: 32
ttccaaagccc ctttattaccacta

<210> SEQ ID NO 33
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: qRT-PCR Probe, TaqMan probe at position 123-142

<400> SEQUENCE: 33
tgactgacctgtaataacca

<210> SEQ ID NO 34
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: qRT-PCR Primers, FWD Primer at position 89-106

<400> SEQUENCE: 34
agggcactgg tgctaatg

<210> SEQ ID NO 35
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: qRT-PCR Primers, Reverse primer at position 146-124

<400> SEQUENCE: 35
tccaaagcct tagcgttaacc atc

<210> SEQ ID NO 36
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: qRT-PCR Probe, TaqMan probe at position 108-122

<400> SEQUENCE: 36
<210> SEQ ID NO 37
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCR Primers, FWD Primer at position 12-35

<400> SEQUENCE: 37
atgcaacct aaaaaacactt tgtg 24

<210> SEQ ID NO 38
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCR Primers, Reverse primer at position 77-97

<400> SEQUENCE: 38
tttctcgga cgattg tgtg c 21

<210> SEQ ID NO 39
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCT Primer, TaqMan probe at position 38-56

<400> SEQUENCE: 39
actaagaaga gaagatt 18

<210> SEQ ID NO 40
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCR Primers, FWD Primer at position 249-270

<400> SEQUENCE: 40
aggtttgcca agggacatta ag 22

<210> SEQ ID NO 41
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCR Primers, Reverse primer at position 314-291

<400> SEQUENCE: 41
acacaccta cttgagatgc aaaa 24

<210> SEQ ID NO 42
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCT Primer, TaqMan probe at position 272-299

<400> SEQUENCE: 42
aggtgctgt aagagaat 18
-continued

<210> SEQ ID NO 43
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCR Primers, FWD Primer at position 42-60

<400> SEQUENCE: 43

cgcacgtctgcgtggtat

<210> SEQ ID NO 44
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCR Primers, Reverse primer at position 107-86

<400> SEQUENCE: 44

gatcgaagctgtggtcata tc

<210> SEQ ID NO 45
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: qRT-PCT Primer, TaqMan probe at position 62-84

<400> SEQUENCE: 45

cgagaaatatatctttgaag gas

<210> SEQ ID NO 46
<211> LENGTH: 544
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 46

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

Pro Arg Ser Ile Tyr His Tyr Gly Asn Lys Lys Gln Leu Leu Leu Lys Thr
 180 185 190
Lys Ser Phe Phe Arg Cln Glu Val Glu Asn Aen Thr Leu Gly Ser Pro
 195 200 205
Ala Ala Ser Glu Leu Leu Glu His Leu Lys Pro Thr Tyr Trp Phe Ser
 210 215 220
Ala His Leu His Val Lys Phe Ala Ala Leu Met Gln His Gln Ala Lye
 225 230 235 240
Asp Lys Gly Gln Thr Ala Arg Ala Thr Lys Phe Leu Ala Leu Asp Lye
 245 250 255
Cys Leu Pro His Arg Asp Phe Leu Gln Ile Leu Glu Ile Glu His Asp
 260 265 270
Pro Ser Ala Pro Asp Tyr Leu Gly Gly Tyr Asp Ile Glu Trp Leu Thr Ile
 275 280 285
Leu Arg Ala Thr Asp Asp Leu Ile Asn Val Thr Gly Arg Leu Trp Asn
 290 295 300 305
Met Pro Glu Asn Asn Gly Leu His Ala Arg Trp Asp Tyr Ser Ala Thr
 310 315 320
Glu Glu Gly Met Lys Glu Val Leu Glu Lys Leu Asn His Asp Leu Lye
 325 330 335
Val Pro Cys Asn Phe Ser Val Thr Ala Ala Cys Tyr Asp Pro Ser Lye
 340 345 350
Pro Gln Thr Gln Met Gln Leu Ile His Arg Ile Asn Pro Gln Thr Thr
 355 360 365 370
Glu Phe Cys Ala Gln Leu Gly Ile Ile Asp Ile Asn Val Arg Leu Gln
 375 380
Lys Ser Lys Glu His His Val Cys Gly Glu Tyr Glu Gln Asp
 385 390 395 400
Asp Val Glu Ser Asn Asp Ser Gly Glu Asp Gln Ser Glu Tyr Asn Thr
 405 410 415
Asp Thr Ser Ala Leu Ser Ser Ile Asn Pro Asp Glu Ile Met Leu Asp
 420 425 430
Glu Glu Glu Asp Glu Asp Ser Ile Val Ser Ala His Ser Gly Met Asn
 435 440 445
Thr Thr Ile Gly Arg Ser Leu Leu Ile Lys Leu Leu Ser Phe Cys Ser
 450 455 460
Phe Ser Asp Val Arg Ile Leu Pro Gly Ser Met Ile Val Ser Ser Asp
 465 470 475 480
Asp Thr Val Asp Ser Thr Ile Asp Arg Glu Gly lys Pro Gly Gly Leu
 485 490 495
Val Glu Ser Gly Asm Gly Glu Asp Leu Thr Val Pro Leu Lys Arg
 500 505 510
Leu Ser Asp Glu His Glu Pro Glu Gly Arg Lys Lye Ile Lys Arg Arg
 515 520 525
Asn Gln Ala Ile Tyr Ala Ala Val Asp Asp Asp Asp Ala Ala
 530 535 540

<210> SEQ ID NO 47
<211> LENGTH: 405
<212> TYPE: PRT
<213> ORGANISM: Saccharomyces cerevisiae
<400> SEQUENCE: 47
<table>
<thead>
<tr>
<th>1</th>
<th>Met Thr Lys Leu Arg Ile Ala Val Gln Gly Cys Cys His Gly Gln Leu</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Asn Gln Ile Tyr Lys Glu Val Ser Arg Ile His Ala Lys Thr Pro Ile</td>
</tr>
<tr>
<td>10</td>
<td>Asp Leu Leu Ile Leu Leu Gly Asp Phe Gin Ser Ile Arg Gly Gin</td>
</tr>
<tr>
<td>15</td>
<td>Asp Phe Lys Ser Ile Ala Ile Pro Pro Lys Tyr Gin Arg Leu Gly Asp</td>
</tr>
<tr>
<td>20</td>
<td>Phe Ile Ser Tyr Tyr Asn Asn Glu Ile Glu Ala Pro Val Pro Thr Ile</td>
</tr>
<tr>
<td>25</td>
<td>Phe Ile Gly Gin Asn His Glu Ser Met Arg His Leu Met Leu Pro</td>
</tr>
<tr>
<td>30</td>
<td>His Gly Gly Tyr Val Ala Asn Ile Phe Tyr Met Gly Tyr Ser Asn</td>
</tr>
<tr>
<td>35</td>
<td>Val Ile Trp Phe Lys Gly Ile Arg Ile Gly Ser Leu Ser Gly Ile Trp</td>
</tr>
<tr>
<td>40</td>
<td>Lys Glu Trp Asp Phe Asn Lys Gin Arg Pro Asp Trp Asn Asp Leu Glu</td>
</tr>
<tr>
<td>45</td>
<td>Aan Asn Asn Trp Lys Ala Asn Ile Aan Leu Tyr His Val Arg Ile</td>
</tr>
<tr>
<td>50</td>
<td>Ser Asp Ile Ala Pro Leu Phe Met Ile Lys His Arg Ile Asp Ile Met</td>
</tr>
<tr>
<td>55</td>
<td>Leu Ser His Asp Trp Pro Asn Gly Val Val Tyr His Gly Asp Thr Lys</td>
</tr>
<tr>
<td>60</td>
<td>His Leu Leu Lys Leu Lys Pro Phe Phe Glu Gin Asp Ile Lys Glu Gly</td>
</tr>
<tr>
<td>65</td>
<td>Lys Leu Gly Ser Pro Val Thr Trp Gin Leu Leu Arg Asp Leu Arg Pro</td>
</tr>
<tr>
<td>70</td>
<td>Gin Trp Trp Leu Ser Ala His Leu His Val Arg Phe Met Ala Ser Ile</td>
</tr>
<tr>
<td>75</td>
<td>Lys His Aan Lys Arg Ser His Glu Pro Pro Asn Lys Ser Thr Ser Lys</td>
</tr>
<tr>
<td>80</td>
<td>Thr Lys Lys Asn Asn Aan Ile Asp Leu Asp Leu Ser Ser Asp Glu</td>
</tr>
<tr>
<td>85</td>
<td>Asp Glu Arg Ser Gly Ile Met Asn Cys Gin Glu Glu Aan Glu Tyr Asp</td>
</tr>
<tr>
<td>90</td>
<td>Ser Lys Tyr Gly Glu Thr Arg Phe Leu Ala Leu Asp Lys Cys Leu Pro</td>
</tr>
<tr>
<td>95</td>
<td>Arg Arg Arg Trp Leu Glu Ile Leu Glu Ile Glu Pro Asp Thr Ser His</td>
</tr>
<tr>
<td>100</td>
<td>Ala Ser Trp Lys Asp Glu Asn His Arg Met Phe Trp Asp Pro Glu Phe</td>
</tr>
<tr>
<td>105</td>
<td>Ile Aan Aan Leu Val Ile Cys Gin Lys Aan Lys Aan Leu Leu Ser Asn</td>
</tr>
<tr>
<td>110</td>
<td>Lys Pro Phe Aan Ser Val Asn Trp Ile Glu Leu Ser Gin Ser Aan Arg</td>
</tr>
<tr>
<td>115</td>
<td>Glu Glu Gly Arg Asp Ile Asp Trp Glu Aan Tyr Ala Ile Pro Ala Tyr</td>
</tr>
<tr>
<td>120</td>
<td>Thr Leu Asp Ile Gin Lys Asp Glu Val Arg Gin Thr Lys Ala Phe Ile</td>
</tr>
<tr>
<td>125</td>
<td>Ser Lys Phe Met Thr</td>
</tr>
<tr>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td>180</td>
<td>185</td>
</tr>
<tr>
<td>195</td>
<td>200</td>
</tr>
<tr>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Arg Ser Ile Ala Ile Asn Lys Val Val Glu Lys Tyr Leu Asn Asp Asn
370 375 380
Arg Gln Asp Phe Thr Glu Leu Asn Thr Lys Gln Ile Leu Ser Asn Pro
385 390 395 400
Gln Gln Phe Pro Leu Val Asn Glu Leu Met Pro Ile Ile Glu Asn Asp
405 410 415
Phe Lys Ser Met Gln Lys Asn Ile Thr Asp Glu Asp Phe Phe Met Val
420 425 430
Pro Glu Asn Phe Gln Thr Ile Ala Pro Thr Asp Glu His Thr Glu
435 440 445
Ser Lys Leu Lys Tyr Tyr Pro Asn Asn Gln Thr Ser Glu Tyr Cys Glu
450 455 460
Lys Phe Gly Ile Pro Lys Leu Val Leu Ser Lys Glu Ser Asp Gln
465 470 475

<210> SEQ ID NO 49
<211> LENGTH: 495
<212> TYPE: PRT
<213> ORGANISM: Candida glabrata

<400> SEQUENCE: 49
Met Asn Glu Arg Lys Leu Arg Ile Ala Val Gin Gly Cys Cys His Gly
1 5 10 15
Glu Leu Asn Lys Val Phe Ala Thr Val Ala Asp Met His Lys Arg Thr
20 25 30
Pro Ile Asp Leu Leu Ile Ile Leu Gly Asp Phe Gln Ser Ile Arg Asp
35 40 45
Asn Ser Asp Phe Gln Ser Ile Ser Ile Pro Pro Lys Tyr Gln Lys Leu
50 55 60
Gly Asp Phe His Ala Tyr Tyr Glu Asn Asp Tyr Tyr Arg Ala Pro Val
65 70 75 80
Phe Thr Ile Val Ile Gly Gly His Glu Ser Met Arg His Leu Met
85 90 95
Gln Leu Pro Tyr Gly Gly Tyr Leu Ala Asn Asn Ile Tyr Tyr Met Gly
100 105 110
Tyr Ser Gly Val Val Trp Phe Lys Gly Phe Arg Ile Ala Ala Leu Ser
115 120 125
Gly Ile Trp Lys Glu Trp Asp Phe Glu Lys Arg Pro Ser Trp Lye
130 135 140
Phe Leu Glu Glu Asn Asn Lys Trp Lys Asp Ser Val Arg Gln Leu Tyr
145 150 155 160
His Ile Arg Lys Asp Asp Val Ala Pro Leu Phe Ala Leu Ser Asp Asn
165 170 175
Ile Asp Ile Cys Leu Ser His Asp Trp Pro Ser Gly Val Val His Tyr
180 185 190
Gly Asn Val Lys Gin Leu Leu Lys Tyr Lys Pro Phe Phe Glu Lys Asp
195 200 205
Ile Lys Ser Gly Lys Leu Gly Asn Pro Ile Ala Trp Lys Leu Leu Thr
210 215 220
Asn Leu Lys Pro Arg Trp Trp Phe Ser Ala His Leu His Val Lys Tyr
225 230 235 240
Glu Ala Glu Ile Thr His Asn Lys Arg Arg Leu Ala Asp Ser Lys Gly
Ala Lys Lys Leu Lys Ser Asn Ser Asp Glu Ile Glu Leu Asn Leu Asp 260 265 270
Asp Glu Ser Ser Leu Asp Leu Ser Cys His Asp Ser Leu Asp Ser 275 280 285
Ala Glu His Thr Arg Phe Leu Ser Leu Asp Lys Cys Met Pro Arg Arg 290 295 300
Lys Trp Leu Glu Ile Val Glu Ile Glu Lys Arg Tyr Asp Ser Ile Pro 305 310 315 320
Gln Gly Leu Asp Cys Asp Lys Met Tyr Trp Asp Pro Ser Tyr Ile Ile 325 330 335
Ala Leu Gln Asn Leu Glu Gln Ser Arg Leu Val Ala Asp Thr Pro 340 345 350
Phe Asn Glu Ile Ile Trp Ser Arg Phe Ser Ser Gly His Ile Asp Asp 355 360 365
Ile Asn Trp His Lys Tyr Glu Ile Pro Lys Tyr Glu Ser Gly Leu Gln 370 375 380
Arg Asp Glu Ala Ser Gln Thr Asp Tyr Phe Leu Ser Lys His Met Leu 385 390 395 400
Ser Lys Gly Ser Arg 405

<210> SEQ ID NO 50
<211> LENGTH: 417
<212> TYPE: PRT
<213> ORGANISM: Candida parapsilosis
<400> SEQUENCE: 50
Met Lys Thr Leu Lys Val Ala Ile Glu Gly Cys Cys His Gly Asp Leu 1 5 10 15
Asn Lys Ile Tyr Lys Gly Ile Pro Ser Ser Thr Glu Leu Leu Leu Ile 20 25 30
Cys Gly Asp Phe Gln Ala Leu Arg Asn Thr Ser Asp Tyr Gln Ala Leu 35 40 45
Ser Val Pro Glu Lys Tyr Arg Asp Leu Gly Asp Phe Gln Ser Tyr Tyr 50 55 60
Thr Ser Lys Lys Ala Pro Val Leu Thr Ile Phe Ile Gly Gly Asn 65 70 75 80
His Glu Ser Ser Ser Tyr Leu Gln Glu Leu Lys Tyr Gly Gly Trp Val 90 95 100
Asa Pro Asn Ile Tyr Tyr Leu Gly Phe Gly Ser Val His Tyr Lys 105 110
Gly Leu Ser Ile Cys Gly Trp Ser Gly Ile Tyr Asn Pro His Thr Tyr 115 120 125
Met Asn Lys Ser Phe Asn Val Glu Arg Leu Pro Phe Asp Ser Asn Ser 130 135 140
Ile Arg Ser Val Tyr His Gln Lys Leu Ser Asn Phe Leu Lys Met Tyr 140 150 155 160
Leu Gln Arg Asp Met Asp Ile Val Leu Ser His Asp Trp Pro Val Gly 165 170 175
Ile Glu Lys Phe Gly Asp Lys Tyr Arg Leu Leu Lys Gln Lys Pro Phe 180 185 190
Phe Thr Gln Asp Ile Lys Lys Gly Gln Leu Gly Ser Pro Leu Asn
<table>
<thead>
<tr>
<th></th>
<th>195</th>
<th>200</th>
<th>205</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Val Leu Leu His His Leu Lys Pro Arg Tyr Trp Phe Ser Gly His Leu</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>His Val Lys Phe Lys Ala Asn Val Val His Asn Ile Ser Lys Pro Lys</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gln Val Lys Asn Ala Asn Glu Ile Leu Leu Asp Met Glu Ser Ser Asp</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glu Ala Ser Asp Gly Glu Asn Gln Pro Gln Lys Lys Met Lys Pro Asn</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gly His Val Val His Asp Thr Gln Phe Leu Ala Leu Asp Lys Tyr Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pro Arg Arg Ser Tyr Phe Glu Val Ile Ann Ile Pro Ile Leu Glu Ann</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn His Pro Ser Val His Asp Gly Leu Tyr Tyr Asp Lys Arg Ala</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ile Ala Ile Asn Arg Val Val Lys Tyr Arg Ile Asp Gln Lys Thr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glu Phe Glu Ser Met Ser Pro Arg Glu Ile Leu Arg Asp Pro Arg Lys</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu Glu Lys Phe Ile Pro Leu Val Ala Lys Glu Ser Glu Glu Ile Ann</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gln Ile Asp Asp Asn Gln Phe Val Ile Pro Lys Ann Phe Glu Val Val</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala Pro Ala Asp Tyr Asp Gly Glu Leu Lys Tyr Tyr Pro Ann Ala Gln</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thr Glu Glu Tyr Cys Arg Lys Phe Gly Ile Pro Gln Gln Glu Asp Tyr Gln</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 51
<211> LENGTH: 746
<212> TYPE: PRT
<213> ORGANISM: Histoplasma capsulatum

<400> SEQUENCE: 51

Met Gln His Ile Pro Ala Ser Pro Ser Arg Asn Leu Arg Val Ala Leu
1      5       10      15
Glu Gly Cys Lys Tyr Lys Gln Ile Gly Asp Phe His Glu Tyr Tyr Ser
20     25       30      35
Gly Ala Arg Val Ala Pro Tyr Leu Thr Ile Phe Val Gly Gly Asn His
35     40       45      50
Glu Ala Ser Asn His Leu Phe Glu Leu Tyr Tyr Gly Gly Trp Val Ala
50     55       60      65
Pro Asn Ile Tyr Tyr Leu Gly Ala Asn Val Ile Arg Cys Gly Pro
65     70       75      80
Leu Arg Ile Ala Gly Ile Ser Gly Ile Trp Lys Gly Tyr Asp Tyr Arg
85     90       95      100
Lys Ser His Phe Glu Arg Leu Pro Tyr Ann Arg Ala Asp Met Glu Ser
100    105      110     115
Ile Tyr His Val Arg Glu Leu Asp Val Arg Lys Leu Leu Gln Ile Arg
115    120      125     130
Thr Glu Val Asp Leu Gly Leu Ser His Asp Trp Pro Gln Gly Ile Glu
130    135      140     145
Trp His Gly Asp Phe Gln Lys Leu Phe Gln Lys Lys Lys Leu Phe Glu
145 150 155 160
Pro Asp Ala Asn Ser Gly Arg Leu Gly Ser Val Ala Ala Arg Tyr Ile
165 170 175
Met Asp Arg Leu Arg Pro Ala Phe Trp Phe Ser Ala His Leu His Cys
180 185 190
Lys Tyr Ala Ala Ser Leu Thr His Gly Asp Tyr Lys Pro Ala Glu Leu
195 200 205
Lys Asn Arg Phe Asn Pro Asn Pro Asn Pro His Gln Pro Gln Gln
210 215 220
Pro Asp Leu Gly Asp Asp Leu Ser Ser Pro Ser Leu Val Thr Asn
225 230 235 240
Glu Gln Pro Glu Val Pro Gly Ser Thr Asn Glu Gly Val Thr Ala Thr
245 250 255
Arg Ser Val Glu Val Gly Ser Asp Ala Ala His Ser Ser Lys Glu
260 265 270
Val Ser Thr Thr Ile Val Asp Thr Ala Met Ser Glu Gly Ile Ile Met
275 280 285
Ser Thr Leu Gly Gly Asp Asp Glu Ala Thr Thr Arg Ala Ala Glu Ser
290 295 300
Ala Lys Asn Ala Pro Gln Pro Gln Pro Ala Glu Gly Ala Arg Asp
305 310 315 320
Arg Ala Gln Leu Ser Ala Trp Gln Asn Phe His Ser Val Ala Thr Lys
325 330 335
Asn Asp Ala Glu Glu Asn Val Arg Leu Met Lys Glu Ala Ala Glu Tyr
340 345 350
Glu Lys Gln Ile Glu Ala Gly Leu Ile Ser Arg Pro Glu Val Asn Tyr
355 360 365
Gln Leu Thr Trp Lys Val Ala Val Lys Asp Asp Asn Leu Gly Arg
370 375 380
Glu Ile Ala Gly Val Ala Lys Ile Gly Tyr Asn Ala Glu Gin Glu Leu
385 390 395 400
Arg Glu Ile Thr Glu Gin Glu Ile Arg Asp Gly Gly Thr Glu Val Lys
405 410 415
Asp Pro Asp Glu Ile Asp Ile Cys Leu Asp Ser Ser Ser Asp Thr Ser
420 425 430
Glu Lys Leu Glu Gin Lys Asp Thr Ile Ser Thr Lys Thr Arg Ser Thr
435 440 445
Asp Lys Met Glu Ile Glu His Ser Ile Ser Gly Lys Ala Glu Lys
450 455 460
Thr Ala Thr Ala Val Asp Ala Ala Glu Ala Ser Gin Ser Gly Asp
465 470 475 480
Ile Pro Lys Glu Ile Arg Asp Gin Leu Pro Ala Ser Phe Arg Lys Pro
485 490 495
Glu Thr Ile Leu Asp Asp Ala Pro Val Phe Glu Ser Thr Leu Pro Glu
500 505 510
Ala Ile Ser Asn Thr Glu Thr Asn Phe Leu Ala Leu Asp Lys Cys Asp
515 520 525
Arg His Arg Gin Phe Ile Glu Leu Val Glu Tyr Pro Ala Ile Ser Ser
530 535 540
Pro Glu Glu Gly Glu Thr Gly Glu Ser Arg Pro Tyr Gln Leu Lys
<table>
<thead>
<tr>
<th>545</th>
<th>550</th>
<th>555</th>
<th>560</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr</td>
<td>Asp</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td>Trp</td>
<td>Leu</td>
<td>Ala</td>
<td>Ile</td>
</tr>
<tr>
<td>Thr</td>
<td>Arg</td>
<td>Ala</td>
<td>Phe</td>
</tr>
<tr>
<td>Ala</td>
<td>Asp</td>
<td>Glu</td>
<td>Leu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>565</th>
<th>570</th>
<th>575</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr</td>
<td>Leu</td>
<td>Gly</td>
</tr>
<tr>
<td>Asp</td>
<td>Pro</td>
<td>Ser</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>Pro</td>
<td>Thr</td>
<td>Asn</td>
</tr>
<tr>
<td>Lys</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>Ala</td>
<td>Arg</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>580</th>
<th>585</th>
<th>590</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Tyr</td>
<td>Lys</td>
</tr>
<tr>
<td>Pro</td>
<td>Ser</td>
<td>Ile</td>
</tr>
<tr>
<td>Leu</td>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>Glu</td>
<td>Gln</td>
<td>Trp</td>
</tr>
<tr>
<td>Val</td>
<td>Glu</td>
<td>Glu</td>
</tr>
<tr>
<td>Asn</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>595</th>
<th>600</th>
<th>605</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>Pro</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>Met</td>
<td>Thr</td>
<td>Ile</td>
</tr>
<tr>
<td>Pro</td>
<td>His</td>
<td>Asn</td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Ser</td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>610</th>
<th>615</th>
<th>620</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Pro</td>
<td>Tyr</td>
</tr>
<tr>
<td>Asp</td>
<td>Pro</td>
<td>Ala</td>
</tr>
<tr>
<td>Val</td>
<td>Pro</td>
<td>Ile</td>
</tr>
<tr>
<td>Thr</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>Thr</td>
<td>Glu</td>
<td>Met</td>
</tr>
<tr>
<td>Pro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Phe</td>
<td>Cys</td>
</tr>
<tr>
<td>Asp</td>
<td>Leu</td>
<td>Ile</td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>645</th>
<th>650</th>
<th>655</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile</td>
<td>Glu</td>
<td>Asn</td>
</tr>
<tr>
<td>Lys</td>
<td>Phe</td>
<td>His</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
<td>Asp</td>
</tr>
<tr>
<td>Glu</td>
<td>Glu</td>
<td>Arg</td>
</tr>
<tr>
<td>Phe</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>Ala</td>
<td>Arg</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>660</th>
<th>665</th>
<th>670</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>Pro</td>
<td>His</td>
<td>Pro</td>
</tr>
<tr>
<td>Glu</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td>Gln</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>His</td>
<td>Gly</td>
<td>Gln</td>
</tr>
<tr>
<td>Arg</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>675</th>
<th>680</th>
<th>685</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td>His</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>Ser</td>
<td>Asp</td>
<td>Arg</td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
<td>Asn</td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Gly</td>
</tr>
<tr>
<td>Arg</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Arg</td>
</tr>
<tr>
<td>Arg</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>Ala</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Arg</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Glu</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Ala</td>
<td>Pro</td>
</tr>
<tr>
<td>Ile</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**<210> SEQ ID NO 52**

**<211> LENGTH: 795**

**<212> TYPE: PRT**

**<213> ORGANISM: Blastomyces dermatitidis**

**<400> SEQUENCE: 52**

<table>
<thead>
<tr>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>Gln</td>
<td>His</td>
<td>Ile</td>
</tr>
<tr>
<td>Pro</td>
<td>Thr</td>
<td>Pro</td>
<td>Ser</td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Arg</td>
<td>Ser</td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Ala</td>
<td>Leu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>Gly</td>
<td>Cys</td>
</tr>
<tr>
<td>Gly</td>
<td>His</td>
<td>Gly</td>
</tr>
<tr>
<td>Leu</td>
<td>Asp</td>
<td>Ile</td>
</tr>
<tr>
<td>Tyr</td>
<td>Thr</td>
<td>Ser</td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>Thr</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Ile</td>
</tr>
<tr>
<td>Lys</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>Trp</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>Tyr</td>
<td>Thr</td>
<td>Ser</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Ile</td>
</tr>
<tr>
<td>Ile</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>50</th>
<th>55</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>Asp</td>
<td>Phe</td>
</tr>
<tr>
<td>Gln</td>
<td>Ala</td>
<td>Val</td>
</tr>
<tr>
<td>Arg</td>
<td>Asp</td>
<td>Ser</td>
</tr>
<tr>
<td>Tyr</td>
<td>Asp</td>
<td>Ser</td>
</tr>
<tr>
<td>Ser</td>
<td>Cys</td>
<td>Met</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Pro</td>
<td>Gln</td>
</tr>
<tr>
<td>Lys</td>
<td>Tyr</td>
<td>Arg</td>
<td>Gln</td>
</tr>
<tr>
<td>Asp</td>
<td>Phe</td>
<td>His</td>
<td>Glu</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyr</td>
<td>Arg</td>
<td>Tyr</td>
</tr>
<tr>
<td>Ile</td>
<td>Gly</td>
<td>Asp</td>
<td>Phe</td>
</tr>
<tr>
<td>His</td>
<td>Gly</td>
<td>Tyr</td>
<td>Asp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>85</th>
<th>90</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>Pro</td>
<td>Tyr</td>
<td>Leu</td>
</tr>
<tr>
<td>Thr</td>
<td>Ile</td>
<td>Phe</td>
</tr>
<tr>
<td>Val</td>
<td>Gly</td>
<td>Gly</td>
</tr>
<tr>
<td>Glu</td>
<td>Asn</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>100</th>
<th>105</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Asn</td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td>Tyr</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>Tyr</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>115</th>
<th>120</th>
<th>125</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Pro</td>
<td>Asn</td>
</tr>
<tr>
<td>Ile</td>
<td>Tyr</td>
<td>Tyr</td>
</tr>
<tr>
<td>Leu</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>Val</td>
<td>Ile</td>
<td>Arg</td>
</tr>
<tr>
<td>Cys</td>
<td>Gly</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>130</th>
<th>135</th>
<th>140</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro</td>
<td>Leu</td>
<td>Arg</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>Ile</td>
<td>Ser</td>
<td>Gly</td>
</tr>
<tr>
<td>Ile</td>
<td>Trp</td>
<td>Lys</td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Arg</td>
</tr>
<tr>
<td>Tyr</td>
<td>Asp</td>
<td>Tyr</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>145</th>
<th>150</th>
<th>155</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Lys</td>
<td>Pro</td>
<td>His</td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Arg</td>
<td>Leu</td>
</tr>
<tr>
<td>Pro</td>
<td>Tyr</td>
<td>Asn</td>
<td>Arg</td>
</tr>
<tr>
<td>Ser</td>
<td>Asp</td>
<td>Ile</td>
<td>Gln</td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
</tr>
</tbody>
</table>
Arg Thr Gln Val Asp Leu Gly Leu Ser His Asp Trp Pro Gln Gly Ile
\[165\]
170 \[175\]
Glu Trp His Gly Asp Phe Gln Leu Phe Gln Lys Pro Leu Phe
\[180\]
185 \[190\]
Glu Pro Asp Ala Asn Ser Gly Arg Leu Gly Ser Val Ala Ala Arg Tyr
\[195\]
200 \[205\]
Val Leu Asp Arg Leu Arg Pro Pro Tyr Trp Phe Ser Ala His Leu His
\[210\]
215 \[220\]
Cys Lys Tyr Thr Ala Asn Leu Ile His Gly Asp Tyr Lys Pro Ala Gly
\[225\]
230 \[235\]
240 \[245\]
Leu Lys Asp Arg Phe Ala Ser Asn Gln Gln Pro His Gln Pro Gln Leu
\[250\]
255 \[260\]
Gly Asp Ala Gly Ala Gly Ser Pro Ser Ile Ile Ala Asp Glu Pro Ala
\[265\]
270 \[275\]
Glu Val Ser Gly Ser Thr Asn Gln Gly Ile Arg Ser Thr Pro Val Val
\[280\]
285 \[290\]
Glu Glu Val Ser Asn Leu Arg Ile Ser Pro Asn Glu Pro Ser Thr
\[295\]
300 \[305\]
Thr Ile Thr Gly Thr Ala Met Arg Glu Glu Thr Ile Leu Ser Thr Pro
\[310\]
315 \[320\]
Gly Gly Glu Thr Glu Pro Thr Thr Leu Ser Thr Thr Gly Ile Glu Ser
\[325\]
330 \[335\]
Ile Pro Glu Gln Gln Asn Pro Thr Gln Arg Ala Arg Gly Arg Ala
\[340\]
345 \[350\]
Gln Leu Ser Ala Trp Gln Asn Phe His Ser Val Ala Thr Lys Asp
\[355\]
360 \[365\]
Ala Glu Glu Gly Ala Arg Leu Met Glu Gln Ala Ala Asp Tyr Glu Lys
\[370\]
375 \[380\]
Gln Val Glu Ala Gly Phe Ile Ser Arg Pro Glu Val Asn Tyr Gln Leu
\[385\]
390 \[395\]
Thr Trp Lys Lys Ile Gly Val Asp Asp Gly Leu Gln Arg Glu Ile
\[400\]
405 \[410\]
Glu Asp Val Ala Lys Ile Gly Tyr Asp Ser Gin Glu Pro Gly Lys Val
\[415\]
420 \[425\]
430 \[435\]
Thr Glu Gln Glu Arg Gly Gly Val Thr Thr Val Arg Asn Thr Asp
\[440\]
445 \[450\]
Glu Ile Asp Ile Ser Leu Asp Ser Ser Ser Glu Thr Ser Glu Lys Leu
\[455\]
460 \[465\]
Glu Gin Glu Asp Thr Thr Ser Ala Ile Pro Lys Asn Thr Asp Val Met
\[470\]
475 \[480\]
Glu Ile Asp Ser Gly Thr Ser Phe Glu Lys Ala Glu Val Ala His Ile
\[485\]
490 \[495\]
Gln Thr Pro Thr Ala Ala Asn Ala Thr Gly Ala Leu Gin Leu Asp
\[500\]
505 \[510\]
Asp Ile Pro Ala Asp Ile Leu Asp Glu Leu Pro Ala Ser Phe Arg Lys
\[515\]
520 \[525\]
Pro Gin Pro Val Pro Gly His Thr Pro Ile Phe Glu Pro Thr Leu Pro
\[530\]
535 \[540\]
Glu Ala Ile Lys Asn Thr Val Thr Glu Phe Leu Ala Leu Asp Lys Cys
\[545\]
550 \[555\]
Glu Lys Gly Ala Leu Gly Ser His Asp Trp Pro Gln Gly Ile
\[560\]
565 \[570\]
Glu Trp His Gly Asp Phe Gln Leu Phe Gln Lys Pro Leu Phe
\[575\]
-continued

Glu Thr Arg Arg Gln Phe Ile Glu Leu Val Glu Tyr Ser Ala Ile Ser 580 585 590
Ser Pro Glu Glu Glu Ile Gly Glu Ser Arg Pro Tyr Gln Leu 595 600 605
Lys Tyr Asp Lys Glu Trp Leu Ala Ile Thr Arg Ala Phe Ala Asp Glu 610 615 620
Leu Ile Leu Gly Asp Pro Asn Ala Ser Val Pro Pro Asn Lys Gly Asp 625 630 635 640
Ala Gly Tyr Lys Pro Asp Ile Leu Ala Ala Ala Gln Trp Val Glu Glu 645 650 655 660
Asn Ile Val Lys Pro Gly Arg Met Thr Ile Pro His Asn Phe Ser Ile 665 670
Thr Ala Pro Val Tyr Asp Pro Ala Ile Ser Ile Met Thr Thr Glu Met 675 680 685
Pro Pro Glu Cys Thr Asn Pro Gln Thr Ala Arg Phe Cys Glu Leu Val 690 695 700
Gly Ile Glu Asn Lys Phe His Ala Ser Asp Glu Arg Phe Ala Arg 705 710 715 720
Ala Asp Ala Gly Pro Arg Pro Glu Pro Pro Gln Ser Arg Tyr Gly Gln 725 730 735
Arg Ser Arg Gly His Gln Asp Gly Ser Val His Ser Phe Gly Arg Gly 740 745 750
Val Val Val Glu Gly Val Glu Ala Gly Glu Val Gly Val Glu Gly 770 775 780
Glu Val Ala Asn Trp Asp Cys Val Leu Arg Phe 785 790 795

<210> SEQ ID NO 53
<211> LENGTH: 785
<212> TYPE: PRT
<213> ORGANISM: Aspergillus fumigatus

<400> SEQUENCE: 53

Met Glu Val Ser Ala Ala Asn Pro Ala Ser Leu Arg Ala Phe Glu
1  5  10  15
Gly Cys Gly His Gly Arg Leu Asp Ile Tyr Asp Ser Val Thr Arg  20  25  30
Ser Ala Thr Arg Arg Gly Trp Asp Gly Val Asp Leu Val Val Ile Gly  35  40  45
Gly Asp Phe Glu Ala Val Arg Asn Asp Ala Ser Gly Ala Cys Met Ser  50  55  60
Val Pro Glu Lys Tyr Lys Ala Ile Gly Asp Phe His Glu Tyr Tyr Ser  65  70  75  80
Gly Lys Thr Ala Pro Tyr Leu Thr Ile Phe Ile Gly Gly Arg His  85  90  95
Glu Ala Ser Asn Tyr Leu Phe Glu Leu Tyr Tyr Gly Trp Val Ala  100 105 110
Pro Asn Ile Tyr Tyr Leu Gly Ala Ala Asn Val Ile Arg Cys Gly Pro 115 120 125
Leu Arg Ile Ala Gly Leu Ser Gly Ile Trp Lys Gly Tyr Asp Tyr Arg 130 135 140
<table>
<thead>
<tr>
<th>Lys</th>
<th>Pro</th>
<th>His</th>
<th>Phe</th>
<th>Glu</th>
<th>Arg</th>
<th>Leu</th>
<th>Pro</th>
<th>Tyr</th>
<th>Asn</th>
<th>Asn</th>
<th>Asp</th>
<th>Asp</th>
<th>Val</th>
<th>Gln</th>
<th>Ser</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Tyr</td>
<td>His</td>
<td>Val</td>
<td>Arg</td>
<td>Glu</td>
<td>Leu</td>
<td>Asp</td>
<td>Val</td>
<td>Arg</td>
<td>Lys</td>
<td>Leu</td>
<td>Leu</td>
<td>Gln</td>
<td>Ile</td>
<td>Arg</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Thr</td>
<td>Gln</td>
<td>Val</td>
<td>Asp</td>
<td>Leu</td>
<td>Gly</td>
<td>Leu</td>
<td>Ser</td>
<td>His</td>
<td>Asp</td>
<td>Trp</td>
<td>Pro</td>
<td>Asn</td>
<td>Arg</td>
<td>Val</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td></td>
<td>185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
<td>Asp</td>
<td>His</td>
<td>Glu</td>
<td>Thr</td>
<td>Leu</td>
<td>Phe</td>
<td>Ala</td>
<td>Lys</td>
<td>Lys</td>
<td>His</td>
<td>Gly</td>
<td>Phe</td>
<td>Arg</td>
<td>195</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>Glu</td>
<td>Asp</td>
<td>Ser</td>
<td>Asn</td>
<td>Asn</td>
<td>Gly</td>
<td>Arg</td>
<td>Leu</td>
<td>Gly</td>
<td>Ser</td>
<td>Ile</td>
<td>Ala</td>
<td>Ala</td>
<td>Arg</td>
<td>Phe</td>
<td>Val</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>Leu</td>
<td>Asp</td>
<td>Arg</td>
<td>Leu</td>
<td>Arg</td>
<td>Pro</td>
<td>Ala</td>
<td>Phe</td>
<td>Trp</td>
<td>Phe</td>
<td>Ser</td>
<td>Ala</td>
<td>His</td>
<td>Leu</td>
<td>His</td>
<td>Val</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Lys</td>
<td>Phe</td>
<td>Asn</td>
<td>Ala</td>
<td>Val</td>
<td>Val</td>
<td>Gln</td>
<td>His</td>
<td>Gly</td>
<td>Asp</td>
<td>Asn</td>
<td>Leu</td>
<td>Gln</td>
<td>Pro</td>
<td>Asp</td>
<td>Ser</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Leu</td>
<td>Gly</td>
<td>Pro</td>
<td>Thr</td>
<td>Arg</td>
<td>His</td>
<td>Ile</td>
<td>Ala</td>
<td>Ser</td>
<td>Ser</td>
<td>Gln</td>
<td>Arg</td>
<td>Thr</td>
<td>Ser</td>
<td>Ser</td>
<td>Ser</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
<td>Thr</td>
<td>Leu</td>
<td>Thr</td>
<td>Ser</td>
<td>Phe</td>
<td>Gly</td>
<td>Met</td>
<td>Asp</td>
<td>Gly</td>
<td>Ala</td>
<td>Val</td>
<td>Thr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>275</td>
<td>280</td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Val</td>
<td>Leu</td>
<td>Gly</td>
<td>Asp</td>
<td>Glu</td>
<td>Asp</td>
<td>Met</td>
<td>Pro</td>
<td>Thr</td>
<td>Glu</td>
<td>Gln</td>
<td>Ala</td>
<td>Gln</td>
<td>Val</td>
<td>290</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Pro</td>
<td>His</td>
<td>Asn</td>
<td>Phe</td>
<td>Ser</td>
<td>Glu</td>
<td>Asn</td>
<td>Gly</td>
<td>His</td>
<td>Ala</td>
<td>Ala</td>
<td>Asn</td>
<td>Thr</td>
<td>Leu</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>Glu</td>
<td>Asp</td>
<td>Glu</td>
<td>Arg</td>
<td>Leu</td>
<td>Glu</td>
<td>Glu</td>
<td>Pro</td>
<td>Arg</td>
<td>Glu</td>
<td>Leu</td>
<td>Pro</td>
<td>Thr</td>
<td>Ala</td>
<td>Gln</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>325</td>
<td>330</td>
</tr>
<tr>
<td>Ala</td>
<td>Thr</td>
<td>Gln</td>
<td>Gln</td>
<td>Ser</td>
<td>Asn</td>
<td>Leu</td>
<td>Val</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Arg</td>
<td>Thr</td>
<td>Ser</td>
<td>Ser</td>
<td>Pro</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Leu</td>
<td>Lys</td>
<td>Arg</td>
<td>Val</td>
<td>His</td>
<td>Asp</td>
<td>Asp</td>
<td>Gln</td>
<td>Ser</td>
<td>Arg</td>
<td>Ile</td>
<td>Ser</td>
<td>Ala</td>
<td>Trp</td>
<td>Asn</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>355</td>
<td>360</td>
</tr>
<tr>
<td>Asn</td>
<td>Phe</td>
<td>His</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Arg</td>
<td>Asp</td>
<td>Glu</td>
<td>Ala</td>
<td>Ala</td>
<td>Glu</td>
<td>Asn</td>
<td>Ser</td>
<td>Ala</td>
<td>Pro</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Arg</td>
<td>Leu</td>
<td>Glu</td>
<td>Glu</td>
<td>Ser</td>
<td>Gln</td>
<td>Asp</td>
<td>Ser</td>
<td>Ala</td>
<td>Ser</td>
<td>Gln</td>
<td>Leu</td>
<td>Pro</td>
<td>His</td>
<td>Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>385</td>
<td>390</td>
</tr>
<tr>
<td>Leu</td>
<td>Thr</td>
<td>Trp</td>
<td>Arg</td>
<td>Lys</td>
<td>Ile</td>
<td>Ser</td>
<td>Val</td>
<td>Asp</td>
<td>Glu</td>
<td>Asp</td>
<td>Pro</td>
<td>Val</td>
<td>Arg</td>
<td>Lys</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>405</td>
<td>410</td>
</tr>
<tr>
<td>Val</td>
<td>Thr</td>
<td>Thr</td>
<td>Val</td>
<td>Glu</td>
<td>Lys</td>
<td>Pro</td>
<td>Ala</td>
<td>Asp</td>
<td>Glu</td>
<td>Asn</td>
<td>Gln</td>
<td>Ser</td>
<td>Glu</td>
<td>Thr</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>Lys</td>
<td>Gln</td>
<td>Lys</td>
<td>Thr</td>
<td>Gly</td>
<td>His</td>
<td>Ala</td>
<td>Val</td>
<td>Ser</td>
<td>Ala</td>
<td>Thr</td>
<td>Lys</td>
<td>Asn</td>
<td>Ser</td>
<td>Asp</td>
<td>Glu</td>
<td>430</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>435</td>
<td>440</td>
</tr>
<tr>
<td>Ile</td>
<td>Pro</td>
<td>Leu</td>
<td>Asp</td>
<td>Leu</td>
<td>Asp</td>
<td>Ser</td>
<td>Asp</td>
<td>Ser</td>
<td>Asp</td>
<td>Glu</td>
<td>Ile</td>
<td>Ser</td>
<td>Thr</td>
<td>Thr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>450</td>
<td>455</td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Thr</td>
<td>Gln</td>
<td>Gly</td>
<td>Ala</td>
<td>Thr</td>
<td>Gln</td>
<td>Lys</td>
<td>Gln</td>
<td>Ala</td>
<td>Ala</td>
<td>Val</td>
<td>Thr</td>
<td>Gln</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>465</td>
<td>470</td>
</tr>
<tr>
<td>Pro</td>
<td>Ala</td>
<td>Ala</td>
<td>Pro</td>
<td>Asp</td>
<td>Val</td>
<td>Thr</td>
<td>Gly</td>
<td>Ser</td>
<td>Lys</td>
<td>Leu</td>
<td>Ser</td>
<td>Gly</td>
<td>Thr</td>
<td>Thr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>485</td>
<td>490</td>
</tr>
<tr>
<td>Gln</td>
<td>Gln</td>
<td>Asp</td>
<td>Lys</td>
<td>Pro</td>
<td>Gln</td>
<td>Val</td>
<td>Arg</td>
<td>Ser</td>
<td>Leu</td>
<td>Asp</td>
<td>Ser</td>
<td>Gln</td>
<td>Asp</td>
<td>Val</td>
<td>Arg</td>
<td>495</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>505</td>
</tr>
<tr>
<td>Asn</td>
<td>Leu</td>
<td>Leu</td>
<td>Pro</td>
<td>Thr</td>
<td>Ser</td>
<td>Phe</td>
<td>Ser</td>
<td>Glu</td>
<td>Pro</td>
<td>Glu</td>
<td>Ser</td>
<td>Phe</td>
<td>Val</td>
<td>Ser</td>
<td>Gln</td>
<td>510</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>515</td>
<td>520</td>
</tr>
<tr>
<td>Asp</td>
<td>Val</td>
<td>Arg</td>
<td>Asn</td>
<td>Gln</td>
<td>Leu</td>
<td>Pro</td>
<td>Ala</td>
<td>Ser</td>
<td>Phe</td>
<td>Ser</td>
<td>Arg</td>
<td>Val</td>
<td>Asp</td>
<td>Cys</td>
<td>Pro</td>
<td>525</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>530</td>
<td>535</td>
</tr>
<tr>
<td>Val</td>
<td>Ser</td>
<td>Gln</td>
<td>Asp</td>
<td>Val</td>
<td>Arg</td>
<td>Asn</td>
<td>Gln</td>
<td>Leu</td>
<td>Pro</td>
<td>Ser</td>
<td>Ser</td>
<td>Phe</td>
<td>Ser</td>
<td>Arg</td>
<td>Pro</td>
<td>540</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>545</td>
<td>550</td>
</tr>
</tbody>
</table>
Gln Ala Thr Pro Lys Leu Asp Pro Ser Val Ser Glu Pro Val Pro Glu
565 570 575
Thr Ile Thr Asn Lys Thr Thr Arg Phe Leu Ala Leu Asp Lys Cys Glu
580 585 590
Pro Lys Arg His Phe Leu Glu Leu Leu Glu Ile Pro Ile Val Ser Glu
595 600 605
Gln Asn Gly Ser Gln Arg Thr Arg Pro Phe Arg Leu Glu Tyr Asp Lys
610 615 620
Glu Trp Leu Ala Ile Thr Arg Val Phe Ala Asp Glu Leu Gln Leu Gly
625 630 635 640
Asp Leu Ala Val Gln Met Gln Pro Asp Arg Gly Gln Ala Phe Tyr Lys
645 650 655
Pro Leu Ile Glu Glu Ala Glu Gln Trp Val Glu Glu Asn Val Val Lys
660 665 670
Ala Gly Lys Met Met Val Pro Glu Asn Phe Thr Pro Thr Ala Pro Phe
675 680 685
Phe Asp Pro Ala Val Pro Ile Thr Thr Asp Glu Leu Pro Pro Glu Phe
690 695 700
Thr Asn Pro Gln Thr Ala Gln Phe Cys Glu Leu Ile Gly Ile Glu Asn
705 710 715 720
Lys Phe His Leu Ser Asp Glu Glu Arg Gln Ala Arg Val Glu Ala Gly
725 730 735
Pro Arg Pro Asn Lys Pro Lys Pro Gln Gly Gly Tyr Trp Asn Arg Gly Arg
740 745 750
Arg Arg Asn Tyr Asn Asn Asn Arg Gly Gly Gly Ser Gin Trp Trp
755 760 765
Gly Arg Gly Ala Gly Arg Asp Arg Gly Arg Ser Gly Gly Asn Gin Arg
770 775 780
Try
785
<210> SEQ ID NO 54
<211> LENGTH: 606
<212> TYPE: PRT
<213> ORGANISM: Cryptococcus neoformans

<400> SEQUENCE: 54
Met Arg Ile Ala Ile Gln Gly Cys Ser His Gly Ser Leu Ala Gln Ile
1     5     10     15
Tyr Asp Val Val Asn Tyr Tyr Ser Gin Thr Lys Asn Pro Ile Asp
20    25    30
Leu Leu Leu Cys Gly Asp Phe Gln Ala Leu Arg Ser Lys His Asp
35    40    45
Tyr Ala Ser Leu Ala Val Pro Ala Lys Phe Lys Gln Leu Gly Ser Phe
50    55    60
His Gln Tyr Tyr Ser Gly Glu Arg Val Ala Pro Val Leu Thr Ile Val
65    70    75    80
Ile Gly Gly Asn His Glu Ala Ser Asn Tyr Met Trp Glu Leu Tyr His
85    90    95
Gly Gly Trp Leu Ala Pro Ser Ile Tyr Tyr Leu Gly Ala Asa Gly Ser
100   105   110
Val Tyr Val Asn Gly Leu Arg Ile Val Gly Ala Ser Gly Ile Tyr Lys
115   120   125
-continued

Gly Phe Asp Tyr Arg Lys Gly His Phe Glu Lys Val Pro Tyr Asn Asp 130 135 140
Lys Glu Leu Arg Ser Ile Tyr His Ile Arg Glu Tyr Asp Val Glu Lys 145 150 155 160
Leu Met His Leu Thr Pro Ser Pro Ser Thr Ile Phe Leu Ser His Asp 165 170 175
Trp Pro Thr Thr Ile Ala His His Gly Asn Lys Asn Ala Leu Leu Lys 180 185 190
Arg Lys Pro Phe Phe Arg Asp Glu Ile Glu Lys Asn Thr Leu Gly Ser 195 200 205
Pro Pro Leu Leu Arg Leu Met Asn His Phe Glu Pro Ser Tyr Trp Phe 210 215 220
Ser Ala His Leu His Val Lys Phe Ala Ala Leu Tyr Glu His Gln Ala 225 230 235 240
Pro Ser His Gly Pro Asp Val Asp Gly Ala Pro Leu Pro Leu Pro 245 250 255
Ala Met Ser Thr Ala Ile Ala Gln Thr Gly Asn Asn Pro Asp Glu Ile 260 265 270
Gln Ile Asp Glu Glu Met Asp Glu Gly Asn Pro Asp Glu Ile Ile Val 275 280 285
Glu Asp Glu Gly Glu Ile Ile Val Arg Pro Arg Glu Val Asn Pro 290 295 300
Asp Glu Ile Val Met Asp Glu Phe Asp Asp Pro Pro Ala 305 310 315 320
Val Pro Gln Pro Leu Pro Ile Thr Thr Ser Ser Val Val Asn Pro Glu 325 330 335
Glu Ile Thr Ile Ser Asp Glu Phe Asp Ala Pro Met Ala Val Ser 340 345 350
Gln Ser Pro Gln Pro Leu Pro Pro Thr Arg Ala Asn Ala Ser Asn Pro 355 360 365
Glu Glu Ile Ala Ile Ser Asp Glu Phe Asp Asp Pro Ala Pro Val 370 375 380
Ala Gln Pro Leu Thr Ala Ile Asp Glu Ser Thr Asp Leu Ile Ala Gln 385 390 395 400
Ser Arg Ser Asn Pro Ser His Pro His Val Ala Gly Thr Ile Ala Pro 405 410 415
Pro Ala Ser Asp Ser Thr Ala Pro Arg Val Met Gln Glu Ala Arg Gln 420 425 430
Glu Gln Gln lys Trp Glu Leu His Gly Gly Lys Gly Met Glu Gly Val 435 440 445
Thr Lys Phe Leu Ala Leu Asp Lys Cys Gly Pro Gly Lys Asp His Met 450 455 460
Gln Phe Leu Glu Ile Pro Asp Pro Ser Pro Ala Ile Pro Gly Pro 465 470 475 480
Pro Arg Leu Thr Tyr Asp Pro Glu Trp Leu Ala Ile Ser Arg Ala Phe 485 490 495
His Pro Tyr Leu Ser Thr Ser Tyr Gln Pro Ile Pro Leu Pro Ser Pro 500 505 510
Asp Val Leu Glu Gln Met Val Lys Asp Glu Val Thr Arg Ile Lys Glu 515 520 525
Glu Gly Leu Leu Val Pro Ala Val Pro Glu Lys Gly Ala Val Glu Gly
Gln Glu Gly Leu Val Trp Glu Lys Gly Lys Val Asp Val Gly Arg Val
Gln Arg Phe Trp Trp Thr Ala Pro Pro Glu Gly His Pro Gly Gly Asn
Asp Ala Ala Trp Tyr Thr Asn Pro Gin Thr Glu Ala Phe Cys Gly Met
Leu Gly Val Gln Asn Lys Ile Asn Pro Pro Val Asn Arg Ser

<210> SEQ ID NO 55
<211> LENGTH: 340
<212> TYPE: PRT
<213> ORGANISM: Rhizopus delemar
<400> SEQUENCE: 55
Met Met His Gln Lys Ile Ala Ile Glu Gly Cys Cys His Gly Glu Leu
1  5  10  15
Asp Lys Ile Tyr Asn Ala Val Arg Glu Glu Glu Ala Arg Tyr Gly Gin
20  25  30
Lys Val Asp Leu Val Leu Ile Cys Gly Asp Phe Gin Ala Leu Arg Asn
35  40  45
Glu Ser Asp Leu Ala Cys Met Ala Val Pro Asp Lys Phe Lys Thr Met
50  55  60
Gly Thr Phe Trp Lys Tyr Ser Gin Ala Arg Ala Pro Tyr Pro
65  70  75  80
Thr Ile Phe Ile Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
95  100 105 110
Leu Tyr His Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
115 120 125
Ala Gly Val Ile Asn Phe Glu Gly Leu Arg Ile Gly Glu Leu Ser Gly
130 135 140
Ile Tyr Lys Gin Asn Asp Tyr His Ile Gly His Gly Thr Val Pro
145 150 155 160
Tyr Asn Asn Ser Gin Met Ser Gin His Tyr His Val Arg Gin Tyr Asp
165 170 175
Val Arg Lys Leu Leu Gln Gin Glu Gin Gin Gin Gin Gin Gin Gin
180 185 190
His Asp Trp Pro Arg Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
195 200 205
Ser Leu Gin Ser Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin
210 215 220
Ala Arg Trp Phe Ala Ala His Gin Gin Gin Gin Gin Gin Gin Gin Gin
225 230 235 240
Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
245 250 255
Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
260 265 270
Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
275 280 285
<table>
<thead>
<tr>
<th></th>
<th>290</th>
<th>295</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>Cys</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Pro</td>
<td>Val</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Pro</td>
<td>Arg</td>
<td>Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 56
<211> LENGTH: 706
<212> TYPE: PRT
<213> ORGANISM: Coccidioides immitis

<400> SEQUENCE: 56

Met Ala Ser Asp Leu Pro Thr Gln Lys Gly Phe Arg Leu Ala Ile Glu
1    5 10 15
Gly Cys Gly His Gly Lys Leu His Glu Ile Tyr Asp Ser Val Lys Lye
20   25 30
Ser Ala Glu Ala Lys Gly Trp Asp Gly Val Asp Leu Val Ile Ile Gly
35   40 45
Gly Asp Phe Gln Ala Val Arg Asn Ser Asn Asp Met Ala Cys Met Ala
50   55 60
Val Pro Ala Lys Tyr Lys Lys Ile Gly Asp Phe His Glu Tyr Tyr Ser
65   70 75 80
Gly Ala Arg Val Ala Pro Tyr Leu Thr Ile Phe Val Gly Gly Asn His
85   90 95
Glu Ala Ser Asn His Leu Phe Glu Leu Tyr Gly Gly Trp Val Ala
100  105 110
Pro Asn Ile Tyr Tyr Leu Gly Ala Ala Asn Ser Val Ile Arg Cys Gly Pro
115  120 125
Leu Arg Ile Ala Gly Met Ser Gly Ile Trp Lys Gly Tyr Asp Tyr Arg
130  135 140
Arg Gln His Phe Glu Arg Leu Pro Tyr Gly Asp Ala Leu Arg Ser
145  150 155 160
Ile Tyr His Val Arg Glu Ile Asp Val Arg Lys Leu Leu Gln Val Arg
165  170 175
Thr Gln Val Asp Ile Gly Ile Ser His Asp Trp Pro Gln Ala Ile Glu
190  195 200
Trp Thr Gly Asp Val Asp Leu Phe Arg Arg Lys Pro His Phe Val
195  200 205
Lys Asp Ala Glu Ser Gly Lys Leu Gly Ser Pro Ala Val Arg Tyr Val
210  215 220
Leu Asp Arg Leu Arg Pro Ala His Trp Phe Ser Ala His Leu His Val
225  230 235 240
Lys Tyr Thr Ser Thr Leu Glu His Lys Ala Tyr Ser Pro Pro Arg Ala
245  250 255
Val Asn Ala His Asn Ile Asp Thr Lys Ser Gin Gin Ser Arg Met Lye
260  265 270
Asp Pro Ala Lys Asp Pro Phe Glu Glu Val Met Ala Lys Pro Met
275  280 285
Gln Ala Cys Val Arg Arg Pro Gin Met Met Thr Pro Gly Ala Ala Ala
290  295 300
Tyr Ser Asp Arg Arg Pro Val Thr Tyr Asn Thr Gin Leu Gin Ser Ser |
Glu Gln Asp Arg Ile Asn Ala Trp Arg Gly Phe Tyr Glu Val Ala Ser
Lys Arg Glu Ala Glu Glu Asn Ala Glu Tyr Leu Lys Ala Ala Asp Glu
Phe Arg Arg Arg Val Asp Ala Gly Glu Ile Glu Lys Pro Lys Ser Asn
Ile Asp Tyr Gln Val Thr Trp Lys Val Val Thr Asp Asp Gly Leu
Ser Arg Glu Val Ser Asp Val Arg Thr Lys Ala Glu Asp Glu Ile
Asn Gln Val Glu Glu Glu Thr Ala Pro Ser Pro Val Lys Asn Ala Asp
Glu Ile Asp Leu Glu Met Glu Ala Ser Glu Thr Ala Glu Thr Pro
Asn Glu Ala Leu Asp Ala Ser Ile Thr Lys Gin Ser Phe Ser Thr Gin
Leu Gln Thr Thr Ala Thr Met Pro Met Pro Pro Ala Glu Phe Asp Gly
Val Ser Asp Glu Leu Arg Glu Gln Leu Pro Ala Ser Phe Gin Lys Arg
Asp Lys Thr Gln Asp Lys Ala Ile Ala Glu Glu Leu Pro Gly Gly
Ile Thr Asn Lys Ala Thr Gin Phe Leu Ala Leu Asp Lys Cys Gin Pro
His Arg Lys Phe Leu Glu Leu Glu Ala Leu Asp Gly Met Val Ser Asn
Asp His Thr Asp Glu Gln Arg Pro Tyr Gin Leu Lys Tyr Asp Asp Glu
Trp Leu Ala Ile Thr Arg Val Phe Ala Glu Gly Phe Val Val Gly Lys
Lys Ser Gin Val Leu Ile Asp Lys Gly Ser Ala Phe Tyr Lys Pro Lys
Ile Ile Asp Ala Glu Ala Trp Val Glu Gin Gin Gin Gin Gin Gin Gin
Lys Met Val Val Pro His Asn Phe Thr Ile Thr Ala Pro Val Tyr Glu
Pro Ser Val Pro Val Thr Pro Glu Gin Pro Phe Gly Tyr Leu Asn
Pro Gin Thr Thr Arg Phe Cys Glu Met Leu Gly Ile Ala Asn Pro Phe
Glu Gin Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
Asp Gin Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
Gly Phe Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
Tyr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
Gly Pro
705
<210> SEQ ID NO 57
<211> LENGTH: 569
<212> TYPE: PRT
<213> ORGANISM: Ustilago maydis

<400> SEQUENCE: 57

Met Lys Leu Ala Ile Gln Gly Cys Ser His Gly Glu Asp Ala Ile
1      5      10      15
Tyr Ala Ser Leu Leu Arg Thr Glu Arg Glu Gin Ser Leu His Ile Asp
20     25     30
Ala Leu Leu Leu Cys Gly Asp Phe Gin Ala Ile Arg Asn His Ser Asp
35     40     45
Leu His Ala Leu Ala Val Pro Gin Lys Tyr Arg Gin Leu Gly Asp Phe
50     55     60
His Ser Tyr Tyr Ser Gly Glu Ile Ala Pro Ile Leu Thr Leu Val
65     70     75     80
Ile Gly Gin His Glu Ala Ser Asn Tyr Met His Glu Leu Tyr His
85     90     95
Gly Gly Trp Leu Ala Pro Asn Ile Tyr Phe Leu Gly Ala Ala Gly Val
100    105    110
Ile Glu Leu Asn Gly Ile Val Val Ala Gly Ile Ser Gly Ile Tyr Lys
115    120    125
Glu Lys Asp Tyr Arg Lys Gly Arg Asp Gin Lys Leu Pro Tyr Asp Ala
130    135    140
Gly Ser Ile Arg Ser Cys Tyr His Thr Arg Glu Phe Asp Val Val Arg
145    150    155    160
Leu Lys Ala Leu Lys Asp Gin Val Gin Val Ile Val Met Ser His Asp
165    170    175
Trp Pro Asn Thr Ile Glu Gin Trp Gly Asn Thr Gin Ala Leu Ile Arg
180    185    190
Lys Lys Pro Phe Phe Lys Glu Glu Ile Glu Ser Arg Thr Leu Gly Ser
195    200    205
Pro Pro Leu Met Gin Leu Leu Gin Cyu Leu Lys Pro Ala Pro Thr Phe
210    215    220
Ser Ala His Leu His Val Lys Phe Ala Ala Leu Phe Arg His Glu Gin
225    230    235    240
Met Asp Pro Ala Ile Gin Glu Pro Ser Ser Thr Thr Ala Ala Asn Thr Asn
245    250    255
Pro Gin Ala Leu Gin Gin Leu Gin Gin Leu Gin Gin Gin Gin Gin Gin
260    265    270
Ser Pro Gin Pro Ala Pro Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
275    280    285
Thr Lys Ser Ala Thr Ala Thr Arg Phe Leu Ala Ala Leu His Lys Cys Leu
290    295    300
Pro Gin Thr Gin Phe Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
305    310    315    320
Ala Glu Leu Gin Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
325    330    335
Pro Pro Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
340    345    350
Phe His Ser His Phe Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
355    360    365
Pro Phe Ser Ala Ser Leu Leu Ala Arg Ile Glu Glu Glu Glu Arg Trp 370 375 380
Ile Glu Gln Asn Leu Val Ser Pro Phe Thr Thr Arg Asn Ser Gly Lys 385 390 395 400
Arg Lys Gln Asp Glu Arg Glu Ser Arg Ala Cys Thr Pro Gln Asp 405 410 415
Glu Arg Glu Thr Glu Gly Val Gly Leu Asp Val His Arg Val Glu Gln 420 425 430 435 440 445
Phe Val Arg Thr Ala Pro Ala Pro Phe Glu Pro Gly Gly Leu Ser Gln
Ala Pro Pro Ala Trp Tyr Thr Asn Pro Gln Thr Glu Ala Phe Cys Arg 450 455 460
Phe Leu Gly Ile Glu Asn Lys Ile Asn Pro Arg Pro Asp Ala Phe Gly 465 470 475 480
Gly Ala Ser Cys Tyr Pro His Pro Gln Glu His Leu Ala Ala Ser Phe 485 490 495 500
His Pro Ala Thr Pro Thr Gly Asp Arg Asp Pro Ser Gln Val His Ser 505 510
Glu Ala Val Ser Asp Pro Asn Ala Leu Ala Ile Asp Met Asp Asp Leu 515 520 525
Asp Ser Asp Cys Ser Asp Ala His Ala Asn Gly Asp Gly Phe Lys His 530 535 540 545
Asp Gly Ser Arg Arg Asp Val Leu Thr Leu Ser Asp Asp Glu Leu 550 555 560 565
His Ala Arg Trp Lys Glu Gly Thr Glu

<210> SEQ ID NO 58
<211> LENGTH: 649
<212> TYPE: PRT
<213> ORGANISM: Magnaporthe oryzae

<400> SEQUENCE: 58
Met Gly Asp Ala Glu Thr Gln Thr Phe Thr Ser Pro Asp Gly Leu Arg 1 5 10 15
Val Ala Val His Gly Cys Gly His Gly Val Leu Asn Ala Ile Tyr Ala 20 25 30
Ala Val Ala Ile Ser Cys Lys Glu Arg Gly Trp Asp Thr Val Asp Leu 35 40 45
Leu Ile Gly Gly Asp Phe Glu Ala Val Arg Asn Ala Ala Asp Leu 50 55 60
Ser Val Met Ser Cys Pro Val Tyr Arg Thr Ile Gly Asp Phe His 65 70 75 80
Glu Tyr Tyr Ser Gly Ser Arg Thr Ala Pro Tyr Leu Thr Ile Phe Ala 85 90 95
Gly Gly Asn His Glu Ala Ala Ser His Ser Trp Glu Leu Phe Tyr Gly 100 105 110
Gly Trp Val Ala Pro Asn Ile Tyr Tyr Leu Gly Pro Ala Asn Val Val 115 120 125
Arg Leu Gly Pro Leu Arg Ile Ala Ala Leu Gly Gly Ile Trp Ala Gly 130 135 140 145 150 155 160
Asn Ile Lys Ser Phe Tyr His Val Arg Glu Met Asp Val Arg Lys Leu
165 170 175
Leu Glu Ile Arg Thr Gln Val Asp Ile Gly Leu Ser His Asp Trp Pro
180 185 190
Arg Ala Val Glu Arg His Gly Glu Gly Ala Leu Phe Arg Lys Lys
195 200 205
Pro Phe Leu Arg Asp Glu Ser Lys Ala Gly Thr Leu Gly Asn Pro Ala
210 215 220
Ala Thr Tyr Val Met Asp Arg Leu Arg Pro Ala Tyr Trp Phe Ala Ser
225 230 235 240
His Met His Cys Lys Phe Ala Ala Leu Lys Val Tyr Thr Asp Glu Pro
245 250 255
Pro Thr Glu Asp Gly Val Glu Ala His Lys Ile Asp Asp Gly Pro
260 265 270
Val Ala Gln Ala Lys Asp Leu Thr Ala Glu Ala Ser Ala Pro Thr Ile
275 280 285
Glu Asn Pro Asp Glu Ile Asp Leu Met Asp Asp Asp Asp Ala
290 295 300
Ala Gly Ala Gly Ala Ala Ala Ala Ser Thr Ser Thr Asp Asp Gly Glu
305 310 315 320
Thr Ala Ala Ala Lys Asp Val Val Ser Glu Asn Thr Ser Asp Gly Lys
325 330 335
Val Val Asn Pro Asp Ala Ile Asp Leu Asp Leu Asp Asp Glu Ala
340 345 350
Gln Asp Thr Ala Pro Gly Ala Pro Gly Gly Gin Pro Glu Gly Asp Gly
355 360 365
Glu Gly Lys Ala Lys Pro Leu Thr Ser Thr Glu Lys Ala Thr Asn Glu Asn
370 375 380
Asn Thr Thr Thr Ala Ala Ser Ser Phe Ile Ser Gin Asp Ile Arg
385 390 395 400
Asn Gln Leu Pro Ala Ser Phe Ala Pro Pro Pro Gin Gin Ala Pro Thr
405 410 415
Glu Ser Arg Ala Lys Arg Thr Pro Gly Gin Pro Val Pro Glu Gly Ile
420 425 430
Thr Asn Lys Glu Val Arg Phe Leu Ala Leu Ser Lys Cys Leu Pro Gly
435 440 445
His Asp Phe Leu Gln Leu Cys Asp Ile Ser Pro Leu Arg Asp Ser Ser
450 455 460
Thr Gly Ser Ser Asp Thr Pro Pro Lys Tyr Arg Leu Glu Tyr Asp
465 470 475 480
Pro Glu Trp Leu Ala Ile Thr Arg Val Phe Ala Ser Glu Leu Ile Ile
485 490 495
Gly Asp Ser Asn Ala Thr Ala Thr Thr Asp Leu Gly Glu Glu His Tyr
500 505 510
Lys Pro Leu Ile Gin Ala Glu Arg Thr Trp Val Glu Glu Asn Ile Val
515 520 525
Ala Lys Asp Leu Ala Ile Pro Glu Asn Phe Val Ile Thr Ala Pro
530 535 540
Pro His Ile Pro Gly Gin Pro Glu Val Pro Glu Gin Pro Asp Glu
545 550 555 560
Tyr Thr Asn Pro Gin Thr Ser Ala Phe Cys Glu Leu Leu Gly Val Lys
565 570 575 580
-continued

Asn Leu Trp Asn Ala Thr Asp Glu Glu Arg Leu Glu Arg Lys Asn Gln 580 585 590
Gly Pro Pro Pro Asp Gln Gly Gly Phe Arg Gly Gly Arg Gly Gly Gly 595 600 605
Gly Gly His Gly Gly Arg Gly Arg Tyr 645 645

<210> SEQ ID NO: 59
<211> LENGTH: 546
<212> TYPE: PRT
<213> ORGANISM: Fusarium graminearum

<400> SEQUENCE: 59
Met Thr Thr Asn Ala Phe Glu Ala Gln Gly Val Arg Val Ala Ile Glu 1 5 10 15
Gly Cys Thr Gln Gly His Gly Thr Leu Asp Ala Ile Tyr Ala Ser Val 20 25 30
Glu Glu Ser Cys Lys Gln Arg Gly Trp Asp Val Asp Ile Leu Ile 35 40 45
Ile Gly Gly Asp Phe Gln Ser Val Arg Asn Ala Glu Asp Leu Ser 50 55 60
Met Ser Cys Pro Val Lys Tyr Arg His Leu Gly Asp Phe Pro Lys Tyr 65 70 75 80
Tyr Ser Gly Glu Arg Lys Ala Pro Tyr Leu Thr Ile Phe Ile Ala Gly 85 90 95
Asn His Glu Ala Ser Ser His Leu Trp Glu Leu Tyr Thr Gly Gly Trp 100 105 110
Val Ala Pro Asn Ile Tyr Tyr Met Gly Ala Ala Asn Ile Leu Arg Phe 115 120 125
Gly Pro Leu Arg Ile Ala Gly Leu Ser Gly Ile Trp Lys Gly Phe Asp 130 135 140
Tyr Arg Lys Pro His His Glu Arg Leu Pro Phe Ser Gly Gly Asp Val 145 150 155 160
Lys Ser Trp Tyr His Val Arg Glu Ile Asp Val Arg Leu Leu Gln 165 170 175
Val Gln Thr Gln Val Asp Val Gly Leu Ser His Asp Trp Pro Arg Ala 180 185 190
Val Glu Leu His Gly Asp His Glu Trp Leu Phe Arg Lys Lys Pro Asp 195 200 205
Phe Arg Asn Glu Ser Arg Asp Gly Thr Leu Gly Ser Val Ala Ala Glu 210 215 220
Tyr Val Met Asp Arg Leu Arg Pro Pro His Trp Phe Ser Ala His Met 225 230 235 240
His Val Lys Phe Ala Ala Ile Lys Thr Tyr Ser Glu Ala Gin Pro Glu 245 250 255
Val Glu Thr Lys Gln Glu Leu Ala Pro Ala Ala Pro Val Pro 260 265 270
Ala Thr Glu Asn Asn Pro Asp Glu Ile Asp Leu Asp Met Asp Asp Glu 275 280 285
-continued

Asp Glu Asp Thr Lys Pro Asn Pro Glu Pro Glu Ala Lys Lys Ser Glu
290 295 300
Pro Glu Val Glu Glu Ala Lys Glu Ala Ser Asn Glu Val Ser Asp Glu
305 310 315 320
Leu Arg Ala Gln Leu Pro Ala Ser Phe Ala Arg Pro Gln Pro Lys Lys
325 330 335
Thr Pro Gly Gln Pro Val Pro Pro Gly Ile Thr Asn Lys Glu Val Arg
340 345 350
Phe Leu Ala Leu Asp Lys Cys Leu Pro Gly Arg His Phe Leu Gln Leu
355 360 365
Cys Asp Leu Gln Pro Phe Asn Pro Glu Thr Ser Ser Glu Tyr Pro Pro
370 375 380
Ala Gln Glu Ser Pro Arg Trp Arg Leu Gln Tyr Asp Pro Glu Trp Leu
385 390 395 400
Ala Ile Thr Arg Val Phe His Asp Ser Leu Val Ile Gly Asp Ser Asn
405 410 415
Ala Gln Ser Pro Pro Asp Leu Gly Glu Glu His Tyr Glu Pro Leu Ile
420 425 430
Lys Lys Glu Arg Glu Trp Val Glu Asp Asn Ile Val Lys Ala Gly Lys
435 440 445
Leu Asp Val Pro Tyr Asn Phe Glu Ile Thr Ala Pro Pro His Val Pro
450 455 460
Gly Gly Pro Glu Ile Ala Ser Glu Glu Pro Ser Glu Tyr Thr Asn Pro
465 470 475 480
Gln Thr Ser Lys Phe Cys Glu Ile Met Glu Leu Ser Asn Ile Thr Asp
485 490 495
Ala Thr Asp Glu Arg Arg Glu Arg Lys Ala Glu Gly Pro Pro Lys
500 505 510
Thr Asp Glu Arg Phe Thr Gly Gly Arg Gly Gly Arg Gly Arg Gly
515 520 525
530 535 540
Arg Trp
545

<210> SEQ ID NO 60
<211> LENGTH: 566
<212> TYPE: PRT
<213> ORGANISM: Plasmodium vivax
<400> SEQUENCE: 60

Met Ile Ile Ala Val Val Gly Cys Thr His Gly Glu Leu Asn Phe Ile
1  5 10  15
Tyr Ala Thr Ile Glu Lys Leu Gln Asp Asn Asn Phe Lys Val Asp
20 25 30
Leu Leu Ile Cys Gly Asp Phe Glu Cys Val Arg Tyr Gly Val Asp
35 40 45
Asn Asp Cys Leu Asn Val Pro Asn Lys Tyr Lys Glu Glu Asn Asp
50 55 60
Phe Arg Asp Tyr Phe Thr Gly Lys Lys Ala Lys Val Leu Thr Ile
65 70 75 80
Phe Ile Gly Gly Asn His Glu Ala Val Asn Val Leu Lys Gln Leu Tyr
85 90 95
Tyr Gly Gly Trp Val Ala Pro Asn Ile Tyr Phe Leu Gly Tyr Ser Asn 100 105 110
Val His Asn Ile Asn Asp Phe Arg Ile Cys Ser Leu Ser Gly Ile Tyr 115 120 125
Lys Lys Tyr Asn Phe Tyr Lys Tyr Asn Gly His Tyr Pro Tyr Asp 130 135 140
Glu Ile Ser Lys Val Ser Ala Tyr His Ile Arg Lys Phe Glu Ile Glu 145 150 155 160
Lys Leu Lys Leu Leu Lys Glu Lys Ile Asp Ile Val Val Thr His Asp 165 170 175
Trp Pro Asn Asn Ile Glu His Gly Asp Val Asn Asp Leu Val Arg 180 185 190
Arg Lys Phe His Phe Gln Ser Asp Ile Tyr Asn Thr Leu Gly Asn 195 200 205
Pro His Thr Glu Phe Leu Asn Leu Lys Pro Tyr Phe Trp Phe 210 215 220
Ser Ser His Leu His Val Lys Tyr Ser Ala Ile Phe Leu His Ser Asp 225 230 235 240
Lys Arg Asn Tyr Thr Arg Phe Leu Ser Leu Asp Lys Ala Glu Pro Arg 245 250 255
Lys His Phe Ile Gln Ile Leu Asn Ile Glu Lys Arg Asn Asn Ile Pro 260 265 270
Tyr Leu Ser Phe Asp His Leu Pro Arg Pro Ser Ala Asn Asp Pro Asp 275 280 285
Gly Lys Ser His Phe Asn Glu Asp Tyr Glu Leu Leu Gln His 290 295 300
Val Glu Asp Val Gln Arg Arg Ala Glu Gly Gly Gly Lys Gly His 305 310 315 320
Ser Gly Ala Ala Ala Gln Ala Lys Glu Asn Ala Pro Val Glu Ala Ala 325 330 335
Thr Arg Glu Ala Ala Thr Arg Glu Ala Val Lys Glu Gln Glu Ala Ala 340 345 350
Val Glu Thr Ser Pro Gly Glu Ala Ala Thr Lys Glu Asp Ala Pro Gly 355 360 365
Glu Ala Asp Pro Gln Glu Pro Pro Pro Gln Glu Asn Ala Ala Pro Glu 370 375 380
Arg Lys Lys Leu Phe Ile Cys Tyr Asp Glu Glu Trp Leu Ala Ile Leu 385 390 395 400
Lys Ala Asn Gln His Leu Val Ser Glu Gly Cys Asp Lys Asp Tyr Asn 405 410 415
Leu Glu Lys Leu Lys Cys Pro Ser Lys Glu Asp Phe Glu Tyr Ile Arg 420 425 430
Asp Lys Leu Lys Glu Leu Lys Thr Ser Val Lys Gly Asp Tyr 435 440 445
Tyr Leu Val His Gly Tyr Asn Thr Pro Ser Tyr His Leu Trp Glu 450 455 460
Gln Arg Glu Leu Phe Leu Ser Arg Phe Asp Phe Glu Leu Arg Met 465 470 475 480
Tyr Asp Asp Phe Glu Arg Leu Phe Phe Ala Glu Glu Val Arg Lys Met 485 490 495
Asp Ala Gly Leu Pro Leu Asp Pro Pro Lys Val Glu Glu Asp Glu Glu
<210> SEQ ID NO 61
<211> LENGTH: 565
<212> TYPE: PRT
<213> ORGANISM: Plasmodium falciparum

<400> SEQUENCE: 61

Met Phe Ile Ala Val Val Gly Cys Thr His Gly Glu Leu Asp Leu Ile
1    5    10  15
Tyr Ser Thr Leu Glu Lys Ile Glu Glu Glu Asp Lys Ile Lys Val Asp
20   25    30
Leu Leu Ile Cys Cys Gly Asp Phe Gin Ser Val Arg Tyr Asn Val Asp
35   40    45
Asn Glu Cys Leu Asn Val Pro Ala Tyr Lys Lys Glu Glu Asn Asp
50   55    60
Phe Val Asp Tyr Phe Thr Gly Lys Lys Ala Ile Thr Thr Ile
65   70    75    80
Phe Val Gly Asp His Glu Ala Met Asn Val Leu Lys Gin Leu Tyr
95   100   105   110
Tyr Gly Gly Trp Val Ala Pro Asn Ile Tyr Tyr Leu Gly Tyr Ser Ser
115  120   125
Val Asp Ile Asn Asn Asn Phe Arg Ile Cys Ser Leu Ser Leu Gly Ile Tyr
130  135  140
Lys Lys Tyr Ser Phe Phe Lys Lys Tyr Tyr Glu Ser Tyr Pro Tyr Thr
145  150  155  160
Asp Ile Thr Lys Val Ser Ala Tyr His Ile Arg Lys Tyr Glu Ile Glu
165
Lys Leu Lys Leu Leu Lys Asn Val Asp Ile Val Val Thr His Asp
170  175
Trp Pro Asn Asn Ile Glu Lys His Gly Asp Val His Asp Leu Leu Arg
180  185   190
Arg Lys Tyr His Phe Gin Ser Asp Val Tyr Asn Thr Leu Gly Asn
195  200   205
Pro His Thr Glu Ile Leu Leu Asn Lys Leu Pro Tyr Phe Trp Phe
210  215   220
Ala Ser His Leu His Val Lys Tyr Ser Ala Leu Tyr Ile His Asn Asp
225  230  235  240
Gln Lys Gln Tyr Thr Arg Phe Leu Ser Leu Asp Lys Ala Gin Glu Tyr
245  250  255
Lys His Phe Ile Gln Ile Leu Asn Ile Thr Lys Lys Asp Ser Ser
260  265  270
Ile Asp Leu Leu Asp His Val Pro Lys Val Leu Leu Pro Glu Pro
275  280  285
Gly Ser Lys Met Asp Ile Gln Asn Asp Ala Gin Pro Asn His Asp Leu
<210> SEQ ID NO 62
<211> LENGTH: 496
<212> TYPE: PRT
<213> ORGANISM: Toxoplasma gondii

<400> SEQUENCE: 62

Met Lys Ile Ala Ile Glu Gly Cys His Gly Glu Leu Asp Ala Ile 1 5 10 15
Tyr Ser Ser Leu Ala Arg Leu Glu Met His Lys Met Lys Val Asp 20 25 30
Leu Leu Ile Cys Cys Gly Asp Phe Gln Cys Val Arg Ser Asn Asp 35 40
Leu Gln Phe Leu Ala Cys Pro Pro Tyr Arg Asp Leu Arg Asp Phe 50 55 60
Pro Ala Tyr Phe Arg Gly Glu Glu Ala Pro Cys Leu Thr Val Phe 65 70 75 80
Val Gly Gly Asn His Glu Ala Pro Thr Val Leu Arg Glu Leu Tyr Tyr
Gly Gly Trp Val Ala Pro Lys Ile Phe Tyr Leu Gly His Ala Gly Val
100 105 110
Val Asn Val Gly Gly Val Arg Ile Ala Gly Leu Ser Gly Ile Phe Lys
115 120 125
Ser Gln Asp Tyr Arg Lys Gly Tyr Phe Glu Arg Pro Pro Tyr Thr Glu
130 135 140
Asp Thr Met Arg Ser Ala Tyr His Val Arg Glu Phe Glu Ile Ala Lys
145 150 155 160
Leu Ser Glu Leu Thr Gly Arg Val Asp Ile Val Val Thr His Asp Trp
165 170 175
Pro Glu Gly Ile Tyr Asp Phe Gly Asp Lys Thr Glu Leu Ile Arg Glu
180 185 190
Lys Pro Phe Leu Glu Lys Asp Ile Gln Ala His Glu Leu Gly Asn Pro
195 200 205
His Ser Met Glu Leu Leu Lys Leu Lys Pro Ala Phe Trp Phe Ala
210 215 220
Ala His Leu His Thr Arg Phe Ala Ala Val Tyr Val His Pro Gly Pro
225 230 235 240
Glu Gly Lys Ala Thr Arg Phe Leu Ala Leu Asp Lys Val Leu Pro Arg
245 250 255
Arg Glu Phe Leu Gln Ile Leu Glu Val Glu Pro Leu Pro Ala Gly
260 265 270
Tyr Val Gln Gln Leu Ser Pro Gly Ile Ser Arg Ser Arg Ser Pro Thr Leu
275 280 285
Cys Tyr Asp Glu Glu Trp Leu Ala Ile Leu Arg Ala Asn Gln Gln Val
290 295 300
Leu Pro Val Ser Arg Phe Pro Gln Ser Cys Leu Val Thr Lys Ala
305 310 315 320
Thr Ala Asp Leu Ala Thr Val Lys Asn Leu Ala Ser Leu Gly
325 330 335
Leu Arg Asn Tyr Arg Glu Thr Ser Ser Pro Lys Arg Leu Ser Asn
340 345 350
Ser Val Gly Ala Ala Ala Ala Asp Ala Arg Gly Ser Asp
355 360 365
Gly Asp Arg Arg Ser Ala Arg Glu Glu Gly Lys Glu Glu Ala
370 375 380
Ala Ala Gly Val Ser Ala Gly Ser Val Gin Arg Thr Asp Val Ala
385 390 395 400
Ala Glu Thr Pro Pro Gin Pro Gin Gly Gly Gin Glu Glu Ser Thr Val
405 410 415
Phe Glu Trp Ile Asn Trp Ala Asp Pro Arg Ala Pro Tyr Thr Glu Leu
420 425 430
Lys Glu Gln Arg Leu Phe Leu Arg Asn Ile Leu Gly Phe Asp Glu
435 440 445
Ala Asp Lys Phe Gly Glu Ala Arg Gin Arg Glu Ala Ala Asp Val
450 455 460
Asp Val Pro Val Asp Trp Thr Ser Gly His Val Asp Pro Gin Arg Thr
465 470 475 480
Thr Glu Val Glu Asp Ile Cys Leu Asp Leu Ser Asp Glu Glu Thr Ala
485 490 495
<210> SEQ ID NO 63
<211> LENGTH: 521
<212> TYPE: PRT
<213> ORGANISM: Trypanosoma brucei

<400> SEQUENCE: 63

Met Ser Ser Leu Val His His Phe Phe Asn Val Val Lys Gly Gly Val Thr
1 5 10 15

Glu Arg Thr Ala Pro Ser Ser Ser Gly Gly Ala Thr Glu Thr Phe Ala
20 25 30

Asn Leu His Val Ala Val Val Gly Cys Cys His Gly Glu Leu Asp Lys
35 40 45

Ile Tyr Leu Ala Cys Ser Asp His Gly Val Ser Ser Gly Lys Lys Ile
50 55 60

Asp Phe Val Ile Cys Ala Gly Asp Phe Gln Ala Leu Arg Arg Glu Glu
45 70 75 80

Asp Leu Lys Cys Met Ala Val Pro Glu Lys Tyr Arg Ser Leu Gly Asp
85 90 95

Phe Val Lys Tyr Tyr Gln Gly Lys Arg Ala Pro Tyr Leu Thr Leu
100 105 110

Phe Val Gly Gly Asn His Glu Cys Ser Asp Trp Leu Ala Glu Glu Ser
115 120 125

Tyr Gly Gly Phe Leu Ala Pro Asn Ile Tyr Tyr Leu Gly His Ser Gly
130 135 140

Val Val Val Asp Gly Cys Ile Thr Val Ala Gly Ile Ser Gly Ile
145 150 155 160

Phe Lys Ala His Asp Tyr Val Arg Pro Tyr Pro Asn Arg Pro Phe His
165 170 175

Val Ser Glu Ala Ser Lys Arg Ser Ala Tyr His Val Arg Arg Ile Glu
180 185 190

Val Glu Lys Leu Arg Ala Phe Val Arg Ala Leu Arg His Met Gln Gln
195 200 205

Trp Gly Arg Lys Trp Gly Ala Gin Ser Val Ser Pro Leu Ala Thr Ala
210 215 220

Ala Asn Ile Ala Asn Pro Ala Gin Lys Val Ser Gin Asp Gly Gly Asn
225 230 235 240

Asp Thr Thr Asn Ser His Ile Thr Leu Pro Pro Val Asp Ile Phe Val
245 250 255

Ser His Asp Thr Pro Thr Gly Val Thr Lys Tyr Gly Asp Glu Glu Gln
260 265 270

Leu Leu Arg Tyr Lys Pro Tyr Phe Arg Glu Asp Ile Arg His Gly Val
275 280 285

Leu Gly Asn Pro His Thr Val Lys Leu Leu Gin Asp Ile Lys Pro Arg
290 295 300

Tyr Trp Ile Ala Ala His Leu His Cys Arg Phe Glu Ala Thr Val Pro
305 310 315 320

His Glu Asn Thr Ser Gly Cys Thr Ala Gly Thr Thr Ser Pro
325 330 335

Val Ala Thr Gin Gin Lys Thr Lys Phe Leu Ala Leu Asp Lys Pro Ala
340 345 350

Lys Gly Lys Gly Phe Ile Asp Phe Ile Asp Val Pro Gly Glu Arg Gly
355 360 365
Ala Val Gly Arg Lys Ser Asp Val Asp Arg Val Val His His Pro Leu
Trp Leu Arg Val Leu Arg Glu Ser His Tyr Leu Ser Ala Asn Asp
Asp Ser Trp Ser Ser Glu Thr Cys Asn Phe Leu Gln Ser Ser Glu Glu
Glu Pro Ile Ser Thr Glu Val Ser Ile Pro Ala His Ser Thr Lys Gln
Leu Leu Gln Ser Leu Gly Leu Pro Pro Ser Pro Ile Gln Gln Ala Gln
Pro Gln Ser Thr Ile Ala Val Ala Gly Gly Gly Ser Gly His His
Arg Pro Val Thr Gly Ser Gly His Ala Lys Leu Asp Asp Lys Ala Gly
Ala Pro Asp Ala Asn Cys Ser Ser Val Ala Thr Pro Ala Asp Trp
Asn Gly Ala Arg Thr Glu Asp Gly Val Asp Ala Gly Asn Asp Leu Pro
Trp Val Glu Asp Ala Val Gly Asp Val

<210> SEQ ID NO 64
<211> LENGTH: 503
<212> TYPE: PRT
<213> ORGANISM: Trypanosoma cruzi

<400> SEQUENCE: 64
Met Cys Phe Val Val Val Val Phe Leu Leu Leu Leu Pro
Trp Val Pro Met Cys Gly Val Val Cys Pro His Tyr Ser Ser Phe Phe
Phe Val Arg Phe Val Phe Tyr Arg Leu Ser Gly Gly Lys Gly Cys
Arg Phe Val Leu Tyr Lys Met Ser Leu Val His Phe Phe His Val
Lys Gly Gly Val Thr Asn Thr Ala Lys Asn Asn Thr Gly Ser Ser
Asp Ser Gly Thr Ala Ala Glu Thr His Val Ala Val Glu Gly Cys
Cys His Gly Glu Leu Asp Arg Ile Tyr Ala Ala Cys Ala Asn Glu
Lys Ala Thr Gly Arg Ile Glu Phe Leu Leu Cys Gly Asp Phe
Glu Ala Val Arg Asp Glu Val Asp Arg Ser Met Ala Val Pro Glu
Lys Tyr Cys Val Leu Gly Asp Phe Leu Ala Tyr His Arg Arg Glu Lys
His Ala Pro Tyr Leu Thr Leu Phe Val Gly Gly Asn His Glu Gly Ser
Asp Trp Leu Ala Thr Glu Cys Tyr Gly Gly Phe Leu Ala Pro Asn Ile
Tyr Tyr Ile Gly His Ser Gly Ala Val Ile Val Asp Asp Cys Val Thr
Val Ala Gly Leu Ser Gly Ile Phe Lys Gly His Asp Tyr Ala Arg Pro
Tyr Pro Gly Arg Pro Phe His Ala Ser Glu Ala Lys Arg Ser Ala
Tyr His Val Arg Ile Glu Val Glu Lys Leu Arg Ala Phe Ser Gln
 Ala Leu Glu Arg Met Arg Gln Pro Ala Ser Ser Pro Met Thr Ala Ser
Met Ala Gly Pro Gly Ala Ser Pro Ser Arg Cys Ala Gly Glu Phe Pro
 His Ile Asp Leu Phe Leu Ser His Asp Trp Pro Ala Gly Ile Thr Lys
Tyr Gly Asp Glu Thr Gln Leu Leu Arg Tyr Lys Pro Phe Phe Glu Glu
Asp Ile Arg His Gly Ala Leu Gly Asn Pro His Thr Met Thr Leu Leu
Arg Ala Val Lys Pro Arg Tyr Trp Leu Ala Ala His Leu His Cys Gln
 Phe Glu Ala Thr Ile Pro His Asp Val Glu Asn Asp Ala Ala Ala
 Ala Gly Val Pro Arg Ala Thr Lys Phe Leu Ala Leu Arg Lys Cys Ser
 Lys Gly Lys Gly Phe Ile Asp Phe Ile Asp Val Arg Val Ser Arg Gly
 Pro His Leu Thr Lys Glu Lys Asn Arg Glu Arg Thr Ala Arg Glu Gln
 Glu Arg Val Val His His Pro Leu Trp Leu Glu Val Leu Arg Glu Thr
 His Gly Phe Leu Thr Ser Asn Asn Glu Trp Ser Ala Gly Ser Cys
 Ala Leu Leu Arg Leu Thr Pro Arg Glu Leu Arg Arg Gly Val Trp
Leu Ala Arg Ser Thr Ala Ser Val Leu Glu Ala Leu Val Leu Pro
 Pro Ala Pro Leu Gln Arg Pro Ser Ala Glu Gly Glu Trp Arg Arg Arg
 Arg Thr His Ala Ser Ala Leu
-continued

Tyr Asp Ser Cys Ser Glu His Glu Arg Gin Thr Gly Lys Arg Ile Asp
Val Leu Leu Cys Cys Gly Asp Phe Gin Ala Val Arg Thr Ala Arg Asp
Met Asp Ser Met Ala Val Pro Asp Lys Tyr Lys Val Leu Gly Asp Phe
His Lys Tyr Tyr Ala Asp Val Ser Gly Ala Phe Thr Gly His Lys Ala
Gln Thr Leu Ala Pro Tyr Leu Thr Ile Phe Val Gly Gly Arg Gin His Glu
Asn Ser Asp Leu Leu Ala Gin Glu Ser Tyr Gly Glu Phe Val Ala Pro
Asn Val Phe Tyr Leu Gly His Ser Ser Val Thr Val Asp Asp Cys
Leu Thr Ile Ala Gly Leu Ser Gly Ile Phe Lys Asp Pro Asp Tyr Asp
Arg Pro Tyr Pro Pro Arg Pro Tyr Ala Val Asn Pro Val Ala Lys Lys
Ser Ala Tyr His Val Arg Ile Glu Val Ala Lys Leu His Ala Tyr
Leu Arg Ala Thr Gin Lys Ile Arg Ser Asn Ser Thr Ile Glu Ala Al5
Lys Thr Thr Ser Ala Thr Ser Pro Ala Ala Ser Pro Pro Met Val Asp
Leu Phe Leu Ser His Asp Trp Pro Val Gly Ile Thr Gly Tyr Gly Asp
Glu Ala Gin Leu Leu Arg Phe Lys Pro Tyr Phe Lys Asp Asp Ile Arg
Arg His Ala Leu Gly Asn Pro Tyr Thr Met Arg Leu Leu Gin Glu Ala
Lys Ala Pro Tyr Trp Phe Ala Ala His Leu His Cys Tyr Phe Glu Ala
Thr Val Glu His Pro Ser Ala Gly Ala Thr Glu Thr Met Ala Ala Thr
Ala Ala Ala Ser Thr Lys Phe Val Ala Leu Asp Gin Cys Ala Lys Gly
His Gly Phe Leu Thr Phe Ile Asp Leu Pro Arg Val Arg Arg Gly Gly
Val Arg Ala Ala Pro Ser Glu Ser His Pro Gin His Thr Ala Thr
Val Leu Gly Thr Ser Arg Ile Arg Arg Asp Gin Val Trp Leu Glu Val
Leu Arg Val Ser His Gin Phe Val Ala Ala Asn Arg Thr Val Glu Ala
Val Ala Ser Tyr Arg Ser Ala Thr Arg Pro Ser Ala Ala Ala Leu Leu
Leu Pro Thr Thr Glu Thr Leu Leu Ala Ala Gin Gin Ser Gin Ser Gin
-continued

Thr Lys Gly Ala Asp Gly Arg Ala Ser Pro Ser Ala Thr Arg Arg Asp
485 490 495
Glu Thr Val Trp Gln Asn Arg Ser Thr Arg Cys Ile Gly Gly Ser
500 505 510
Leu Gln Pro His His Pro Arg Ala Arg Thr Glu Ala Thr Arg Ala Ser
515 520 525
Ser Val Ser Thr Ala Ala Pro Lys Ser Ser Thr Pro Leu Trp Tyr Thr
530 535 540
Ala Gly Thr Gln Pro Leu Gln Gin Pro Pro Thr Ser Ala Leu Arg Ile
545 550 555 560
Phe Glu Asp Val Gly Pro Thr Gly Cys Ser Ser Ala Pro Ser Ser Thr
565 570 575
Ser Gly Met Val Ala Gly His Val Ser Ser Ser Phe Ala Cys Thr Asp
580 585 590
Gly Asp Gly Ala Pro Arg Glu Pro Ala Thr Thr Leu Ser
595 600 605
Trp Phe Glu Asp Thr Thr Gin Gin Gin Gin Gin Ser
610 615 620

<210> SEQ ID NO 66
<211> LENGTH: 369
<212> TYPE: PRT
<213> ORGANISM: Cryptosporidium parvum

<400> SEQUENCE: 66
Val Ala Val Ile Gly Cys Cys His Gly Glu Leu Asn Arg Leu Tyr Met
1  5  10  15
Glu Val Glu Lys Tyr Glu Asn Glu Lys Asn Glu Lys Val Asp Leu Ile
20  25
Leu Cys Cys Gly Asp Met Gin Thr Ile Arg Asp Glu Asn Arg Leu Gin
30  35  40  45
Asp Met Ala Val Lys Ser His Arg Ser Lys Gly Asp Phe Trp Glu
50  55  60
Tyr Tyr Glu Gly Leu Lys Lys Ala Pro Lys Val Thr Ile Phe Ile Gly
65  70  75  80
Gly Asn His Glu Thr Pro Asn Val Leu Ile Pro Leu Tyr Tyr Gly Gly
85  90  95
Trp Val Ala Pro Asn Ile Phe Tyr Leu Gly Ser Ser Gly Val Ile Arg
100 105 110
Val Gly Asp Val Arg Val Ala Gly Ile Ser Gly Ile Tyr Lys Asn Tyr
115 120 125
Asp His Phe Arg Gly Tyr Tyr Glu Ser Lys Pro Phe Thr Glu Glu Ser
130 135 140
Lys Arg Ser Trp Tyr His Ile Arg Trp Leu Glu Ile Gin Lys Leu Leu
145 150 155 160
Leu Ile Glu Asn Ile Lys Ser Asn Phe Leu Gly Ser Thr Glu Ser Arg
165 170 175
Lys Val Asp Val Met Ile Ser His Asp Trp Pro Asn Gly Ile Glu Arg
180 185 190
Phe Gly Asn Leu Gly Tyr Leu Ile Arg Arg Lys Pro Tyr Leu Lys Glu
195 200 205
Asp Ile Glu Leu Gly Arg Leu Gly Ile Pro Gly Cys Ile Glu Leu Ile
210 215 220
What is claimed is:

1. An isolated debranching enzyme comprising an amino acid sequence having at least 35% sequence identity to a sequence selected from the group consisting of SEQ ID NO: 46-66.

2. The isolated debranching enzyme of claim 1, wherein the sequence identity is selected from the group consisting of about 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more.

3. The isolated debranching enzyme of claim 1, wherein the amino acid sequence has at least 75% sequence identity to the metallophosphatase domain of a sequence selected from the group consisting of SEQ ID NO: 46-66.

4. The isolated debranching enzyme of claim 3, wherein the sequence identity is selected from the group consisting of about 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more.

5. A method of enriching an RNA population for lariat RNA comprising:
   a. providing an RNA population; and,
   b. contacting the RNA population with a linear RNA degrading enzyme to form a lariat RNA enriched population.

6. The method of claim 5 further comprising contacting the RNA population with a debranching enzyme.

7. The method of claim 6, wherein the debranching enzyme comprises an amino acid sequence having at least 35% sequence identity to a sequence selected from the group consisting of SEQ ID NO: 46-66.

8. The method of claim 5, wherein the linear RNA degrading enzyme is selected from the group consisting of 3' exonuclease, 5' exonuclease, 5'3' exonuclease, or combinations thereof.

9. A method of analyzing lariat RNA in an RNA sample comprising the steps of:
   a. providing an RNA population;
   b. contacting the RNA population with a linear RNA degrading enzyme to form a lariat RNA enriched population; and,
   c. creating a cDNA library from the lariat RNA population.

10. The method of claim 9, wherein the linear RNA degrading enzyme is selected from the group consisting of 3' exonuclease, 5' exonuclease, 5'3' exonuclease, or combinations thereof.

11. The method of claim 9 further comprising contacting the lariat RNA enriched population with a debranching enzyme.

12. The method of claim 10, wherein the debranching enzyme comprises an amino acid sequence having at least 35% sequence identity to a sequence selected from the group consisting of SEQ ID NO: 46-66.

13. The method of claim 9 further comprising sequencing the cDNA library.

14. A kit comprising:
   a. a linear RNA degrading enzyme; 
   b. buffer; and, 
   c. instructions.

15. The kit of claim 14 further comprising a debranching enzyme.

16. The kit of claim 14 further comprising a debranching enzyme buffer.

17. The kit of claim 15, wherein the debranching enzyme comprises a amino acid sequence having at least 35% sequence identity to a sequence selected from the group consisting of SEQ ID NO: 46-66.

18. The kit of claim 17, wherein the sequence identity is selected from the group consisting of about 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% or more.

19. The kit of claim 14, wherein the linear RNA degrading enzyme is selected from the group consisting of 3' exonuclease, 5' exonuclease, 5'3' exonuclease, or combinations thereof.

20. The kit of claim 14 further comprising a 5' decapping enzyme.